

Los marcadores del discurso en el corpus
C-Oral-Rom: anotación pragmática, estrategias
computacionales de etiquetado y aplicaciones a
otros campos

Ana González Ledesma

Director de tesis
Antonio Moreno Sandoval

Departamento de Lingüística General, Lenguas Modernas,
Lógica y Filosofía de la Ciencia,
Teoría de la Literatura y Literatura Comparada

Universidad Autónoma de Madrid

Septiembre 2009

Resumen

En este trabajo presentamos un esquema de anotación pragmática para corpus (Pragmatext), un etiquetador semi-automático de marcadores del discurso (DiMaSTagger) y, finalmente, una aplicación web (Alter Alia) para lingüistas, informáticos y hablantes no nativos del español, a través de la cual se pueden consultar ejemplos de los marcadores del discurso del corpus oral del español C-ORAL-ROM y acceder a sus frecuencias según el género discursivo. El modelo de anotación pragmática con el que se han etiquetado los marcadores es un modelo basado en rasgos, donde cada rasgo hace referencia a uno de los siguientes fenómenos semántico-pragmáticos: modalización discursiva, deixis, actos de habla, evidencialidad y relaciones discursivas. Este trabajo se inserta dentro de un discurso interdisciplinar en el que se relacionan las siguientes parcelas de conocimiento: Marcadores del discurso, Pragmática, Lingüística Cognitiva, Lingüística Computacional e Inteligencia Artificial.

Palabras clave: etiquetador de marcadores del discurso, anotación pragmática, corpus orales, Lingüística Computacional, Enseñanza de Español a Extranjeros y sistemas de diálogo hombre-máquina.

Abstract

In this work, the discourse markers from the Spanish spoken corpus C-ORAL-ROM are been annotated with the pragmatic annotation model PRAGMATEXT. This anotation model is based on a system of features, where each feature corresponds with a semantic-pragmatic phenomenon: discursive modalization, evidentiality, discourse relations, deixis and speech acts. This features will be activated in according on the meaning of the discourse markers. Once the pragmatic annotation model is designed, the corpus is annotated by DiMaSTagger, a discourse markers semi-automatic tagger. And, finally, after all corpus is tagged, we have programmed INTER ALIA, a web interfaz to consult the C-Oral-Rom corpus with an different aplications. For example, you can search a discourse marker in the corpus according to the type of discursive genre, and the web interfaz provides the frequency of use. This work is an interdisciplinary reflection where these areas of knowledge are connected: Discourse Markers, Pragmatics, Cognitive Linguistics, Computational Linguistics and Artificial Intelligence.

Keywords: discourse markers tagger, pragmatic annotation, spoken corpus, computational linguistic, ELE2 and man-machine dialogue system.

Agradecimientos

A Antonio Moreno Sandoval debo agradecerle esa primera oportunidad, hace ya diez años, para trabajar en lo que más me gusta, investigar.

Mi más sincero agradecimiento a La tertulia, conformada por los siguientes investigadores y amigos para toda la vida:

Alicia Mezquita Amezcua, Universidad de Guadalajara.
Daniel Israel, Universidad de Mendoza.
Ana María Díaz Villa, Real Academia de la Lengua.
Doaa Saamy, Universidad del Cairo.
Yazmin Hosny, Universidad del Cairo.
Steven Bermúdez, Universidad de Zulia (Venezuela).
Ana Martínez, Universidad Autónoma de Madrid.
Tania Ramírez, Comisión de Derechos Humanos de México D. F.
Pablo Álvarez Watkins, Facultad de Ingeniería de la UNAM.

A Fernando Sánchez León, a quien le agradezco las aportaciones con las que desinteresadamente ha contribuido a esta tesis.

Una mención especial se merecen Emilia Apter, a la que conocí en mi estancia en Londres, y Sarine Kouyoumdjian, de mi estancia en Florencia. Gracias por vuestro increíble valor humano, vuestra inquietud por el pensamiento metalingüístico y vuestra adicción al análisis y a la conversación. A vosotras os debo mis conocimientos de inglés e italiano y las primeras reflexiones de esta tesis en otras lenguas.

A Concepción Bueno Díaz Munio, gracias por la aplicación de tu inteligencia y tus conocimientos de informática a la resolución de los problemas computacionales que planteaba esta tesis que, sin tu ayuda, nunca habría terminado a tiempo.

Al resto de mis queridos amigos Ignacio e Idoya Bustos, Marta Garrote Salazar, Miguel Pérez Milans, Gustavo Cilleros, Guido, Marga, Pablo, Diego, Isabel, Fanny, Elena López García, Claudia, Isabella, Federica y Linda.

A Alex Formisano, gracias por tus conocimientos de Física.

Al escritor de la novela *El club de los cisnes negros*, Miguel Ángel Salas Roselló.

A mi familia.

*Dedicado a Fernando Sánchez León,
poeta de la Lingüística Computacional.*

*en el mundo, en conclusión,
todos sueñan lo que son,
aunque ninguno lo entiende.*

La Vida es Sueño
Calderón de la Barca
(1635)

Índice general

I	De la teoría	19
1.	Cognición y comunicación, un marco para la dimensión pragmática de la lengua	21
1.1.	Introducción	21
1.2.	La construcción de la subjetividad: el observador de metáforas	23
1.2.1.	Pensamiento, lenguaje y acción: acciones codificadas y acciones inferidas	26
1.2.2.	Verdad, certeza y evidencialidad.	27
1.2.3.	La lógica clásica y el sentido común: conectores lógicos y marcadores del discurso	29
1.2.4.	La codificación lingüística de los estados emocionales	31
1.3.	La construcción de la intersubjetividad: los observadores observados	32
1.4.	La formalización de la intersubjetividad: Formalismo vs. Funcionalismo y Pragmática Computacional	37
2.	Teoría pragmática	41
2.1.	Introducción	41
2.2.	Introducción a la Pragmática	41
2.2.1.	Raíces filosóficas del movimiento	41
2.2.2.	Definiciones del término Pragmática	42
2.2.3.	Relación entre Gramática y Pragmática, Semántica y Pragmática, y Pragmática y otras ciencias	43
2.2.4.	Conceptos de Pragmática: enunciación y enunciado, inferencia, presuposición, implicatura y contexto	45
2.2.5.	El concepto de contexto	49
2.3.	Pragmática y cognición: La teoría de la Relevancia	51
2.4.	Teoría de actos de habla	54
2.4.1.	Conceptos Básicos	54
2.5.	El contrato comunicativo: Grice, Levinson, Lakoff	58
2.6.	La vertiente social de la comunicación	59
2.6.1.	Los seguidores de P. Grice	59
2.6.2.	Un modelo universal de cortesía	60
2.7.	Teoría de la Argumentación de Anscombe y Ducrot	63
2.8.	Conclusiones finales	66
3.	Los marcadores discursivos en las teorías pragmáticas	69
3.1.	Introducción	69
3.2.	Definición de marcador discursivo y características gramaticales	69
3.3.	¿Qué ha dicho la Pragmática de los marcadores discursivos?	72
3.3.1.	Los marcadores discursivos: relevancia y argumentación	72
3.3.2.	Clasificación de los marcadores del discurso	73
3.3.3.	La Estructura textual y los marcadores del discurso	77

3.4. Los marcadores discursivos, Lingüística Computacional e Inteligencia Artificial	81
3.5. Relaciones de coherencia, relaciones retóricas y marcadores del discurso	83
3.6. Final de la parte teórica	84

II De la descripción 87

4. Pragmatest, un modelo de anotación pragmática para corpus	88
4.1. Introducción	88
4.2. Concepción teórica del modelo de anotación pragmática	88
4.3. Descripción de los fenómenos semántico-pragmáticos de PRAGMATEST	91
4.3.1. La codificación verbal de las emociones	91
4.3.2. Las operaciones argumentativas	95
4.3.3. La modalización discursiva	105
4.3.4. La evidencialidad	106
4.3.5. Deixis	108
4.3.6. Actos de habla	109
4.3.7. Metáfora y Unidades Fraseológicas	110
4.4. Conclusión	113
5. Los marcadores discursivos de C-ORAL-ROM clasificados con PRAGMATEST	115
5.1. Introducción	115
5.2. Pragmatest y los marcadores discursivos	115
5.3. Etiquetado en XML	117
5.4. Los marcadores del discurso de C-ORAL-ROM clasificados con PRAGMATEST	121
5.4.1. Marcadores de Operaciones Discursivas	123
5.4.2. Marcadores de Evaluación Emocional	142
5.4.3. Marcadores de Modalización discursiva	153
5.4.4. Marcadores de Evidencialidad	178
5.4.5. Marcadores de Actos de Habla	181
5.4.6. Marcadores de Deixis	186

III Del procesamiento 201

6. Lingüística de corpus y Anotación pragmática: C-ORAL-ROM y Pragmatest	203
6.1. Introducción	203
6.2. C-ORAL-ROM en la tradición lingüística de corpus orales	203
6.2.1. El corpus oral del español C-ORAL-ROM: diseño y características	204
6.2.2. Las cabeceras de las transcripciones	204
6.2.3. La anotación del corpus C-ORAL-ROM: niveles de anotación y formatos	210
6.3. Corpus anotados a nivel semántico-pragmático	212
6.3.1. Pragmatest en C-ORAL-ROM en formato XML	214
7. Estrategias de reconocimiento y clasificación semiautomática de los marcadores del discurso del corpus C-ORAL-ROM: el anotador DiMaStagger y su evaluación	223
7.1. Introducción	223
7.2. Fase de preanotación	224

7.2.1.	Elaboración del lexicón de marcadores del discurso	224
7.2.2.	Reconocimiento de los candidatos a MD en el corpus	224
7.3.	Fase de anotación	227
7.3.1.	Funcionamiento a tiempo real de DiMaStagger	234
7.4.	Fase de evaluación de la anotación	239
7.4.1.	Resultados del grado de acierto de las reglas contextuales	239
7.4.2.	Resultados del grado de acierto sobre CAT vs. MD	245
7.5.	Conclusiones	253

IV De la aplicación 255

8.	Los marcadores del discurso en C-ORAL-ROM, entre la oralidad y la escritura: un análisis de corte cuantitativo	257
8.1.	Introducción	257
8.2.	Oralidad y escritura en los marcadores del discurso de C-Oral-Rom	258
8.3.	Inter Alia para lingüistas	262
8.4.	Frecuencia de marcadores del discurso en función de los géneros discursivos del corpus C-ORAL-ROM	266
8.4.1.	Frecuencias de marcadores de Relación Discursiva	266
8.4.2.	Frecuencias de marcadores de Evaluación Discursiva	273
8.4.3.	Frecuencias de marcadores de Modalización discursiva	277
8.4.4.	Frecuencia de marcadores de Evidencialidad	284
8.4.5.	Frecuencia de marcadores de Actos de Habla	285
8.4.6.	Frecuencia de marcadores de Deixis	287
8.5.	Conclusiones	295
9.	La enseñanza por ordenador de los marcadores del discurso a alumnos italianos con C-ORAL-ROM, un corpus del español oral	297
9.1.	Introducción	297
9.2.	Estado de la cuestión: ELE, Pragmática, ordenadores y corpus	298
9.2.1.	Marcos teóricos y metodológicos para ELE	298
9.2.2.	La enseñanza de la Pragmática y de los marcadores del discurso	301
9.2.3.	La enseñanza de lenguas y el ordenador	303
9.2.4.	La enseñanza de segundas lenguas con corpus	304
9.3.	INTER ALIA, un interfaz Web para el aprendizaje de la pragmática	305
9.3.1.	Declaración de principios	305
9.3.2.	El interfaz Web: Inter Alia	310
9.4.	Conclusiones	315

V De la interpretación 319

10.	Inteligencia artificial y Pragmática: perspectivas de investigación	321
11.	Conclusiones finales y trabajo futuro	329
11.1.	Resumen y novedad del trabajo	329
11.2.	Puntos críticos y trabajo futuro	331
11.3.	Las voces de la razón: retos para la relación entre mente y lenguaje.	333

VI De las fuentes de conocimiento	337
Bibliografía	338

Índice de figuras

7.1. Ejemplo del anotador en la fase de desambiguación categorial	236
7.2. Ejemplo del anotador en la fase de desambiguación de etiquetas	238
8.1. Web para lingüistas: haz click en consultar la frecuencia de un MD en el corpus	263
8.2. Web para lingüistas: introduce un marcador en el cajetín	264
8.3. Web para lingüistas: resultados de frecuencias según el género discursivo .	265
9.1. Primera página de Inter Alia	311
9.2. Segunda página de Inter Alia	312
9.3. Tercera página de Inter Alia	313
9.4. Cuarta página de Inter Alia	314

Índice de tablas

2.1. Fórmula para calcular el riesgo social de la Teoría de la Cortesía	62
2.2. Explicación de las firmas de la fórmula de arriba	62
3.1. Clasificación de marcadores del discurso de J. Portolés.	75
4.1. Concepción teórica del modelo de anotación pragmática para corpus Prag- matext	90
5.1. Tabla de firmas de las etiquetas	119
5.2. Tabla de valores de cada atributo	120
5.3. Relaciones Discursivas Generalization y Concretness	124
5.4. Relaciones Discursivas Coargumentation1, 2, 3	125
5.4. Relaciones Discursivas Coargumentation1, 2, 3	126
5.4. Relaciones Discursivas Coargumentation1, 2, 3	127
5.5. Relaciones Discursivas Contraargumentation	128
5.6. Relaciones Discursivas Anaphoric Concessiveness Cataphoric Conclusion y Anaphoric-Cause-Cataphoric-Consequence	129
5.6. Relaciones Discursivas Anaphoric Concessiveness Cataphoric Conclusion y Anaphoric-Cause-Cataphoric-Consequence	130
5.7. Relaciones Discursivas Option	131
5.8. Relaciones Discursivas Hypothesis	132
5.9. Relaciones Discursivas Cause y CathaphoricCause	133
5.10. Relaciones Discursivas Consequence	134
5.11. Relaciones Discursivas ConclusionSummary	135
5.12. Relaciones Discursivas Reformulation	136
5.13. Relaciones Discursivas: Defined End e Indefinite Transitional End	137
5.13. Relaciones Discursivas: Defined End e Indefinite Transitional End	138
5.13. Relaciones Discursivas: Defined End e Indefinite Transitional End	139
5.14. Relaciones Discursivas: Framework, Discourse Production, Digression, Scalar Meaning	140
5.14. Relaciones Discursivas: Framework, Discourse Production, Digression, Scalar Meaning	141
5.15. Marcadores de significado emocional, Positive Evaluation	143
5.15. Marcadores de significado emocional, Positive Evaluation	144
5.15. Marcadores de significado emocional, Positive Evaluation	145
5.15. Marcadores de significado emocional, Positive Evaluation	146
5.16. Marcadores de significado emocional Negative Evaluation	147
5.16. Marcadores de significado emocional Negative Evaluation	148
5.16. Marcadores de significado emocional Negative Evaluation	149
5.16. Marcadores de significado emocional Negative Evaluation	150
5.17. Marcadores de significado emocional Surprise	151
5.17. Marcadores de significado emocional Surprise	152

5.18. Marcadores de modalizacion: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster	154
5.18. Marcadores de modalizacion: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster	155
5.18. Marcadores de modalizacion: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster	156
5.18. Marcadores de modalizacion: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster	157
5.18. Marcadores de modalizacion: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster	158
5.18. Marcadores de modalizacion: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster	159
5.18. Marcadores de modalizacion: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster	160
5.18. Marcadores de modalizacion: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster	161
5.18. Marcadores de modalizacion: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster	162
5.18. Marcadores de modalizacion: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster	163
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	164
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	165
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	166
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	167
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	168
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	169
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	170
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	171
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	172
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	173
5.19. Marcadores de modalizacion: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge	174
5.20. Marcadores de modalizacion: Interaction	175
5.20. Marcadores de modalizacion: Interaction	176
5.20. Marcadores de modalizacion: Interaction	177
5.21. Marcadores de evidencialidad: Cultural Knowledge, inference, others, first person	179
5.21. Marcadores de evidencialidad: Cultural Knowledge, inference, others, first person	180
5.22. Marcadores de actos de habla: Commitment, wish, rejection, acceptation, ask, compliment, assent, reciprocal action, interaction close, risk, justification, complaint, resignation, greeting open, refuse, greeting close, gratitude, action, warning	182

5.22. Marcadores de actos de habla: Commitment, wish, rejection, acceptance, ask, compliment, assent, reciprocal action, interaction close, risk, justification, complaint, resignation, greeting open, refuse, greeting close, gratitude, action, warning	183
5.22. Marcadores de actos de habla: Commitment, wish, rejection, acceptance, ask, compliment, assent, reciprocal action, interaction close, risk, justification, complaint, resignation, greeting open, refuse, greeting close, gratitude, action, warning	184
5.22. Marcadores de actos de habla: Commitment, wish, rejection, acceptance, ask, compliment, assent, reciprocal action, interaction close, risk, justification, complaint, resignation, greeting open, refuse, greeting close, gratitude, action, warning	185
5.23. Marcadores de deixis: Social Interlocutor	187
5.23. Marcadores de deixis: Social Interlocutor	188
5.23. Marcadores de deixis: Social Interlocutor	189
5.23. Marcadores de deixis: Social Interlocutor	190
5.23. Marcadores de deixis: Social Interlocutor	191
5.23. Marcadores de deixis: Social Interlocutor	192
5.23. Marcadores de deixis: Social Interlocutor	193
5.24. Marcadores de deixis: Textual Contextual	194
5.24. Marcadores de deixis: Textual Contextual	195
5.24. Marcadores de deixis: Textual Contextual	196
5.24. Marcadores de deixis: Textual Contextual	197
5.24. Marcadores de deixis: Textual Contextual	198
5.24. Marcadores de deixis: Textual Contextual	199
6.1. Distribución de la parte informal de C-Oral-Rom	206
6.3. Distribución de la parte formal de C-Oral-Rom	206
6.4. Tabla de signos de transcripción	207
6.4. Tabla de signos de transcripción	208
6.4. Tabla de signos de transcripción	209
7.1. Clasificación de las etapas del etiquetado por dificultad	228
7.2. Separación del programa DiMaSTagger en módulos de dificultad según el tipo el tipo de marcador	230
7.3. Tabla de especificación de los tipos contextuales, donde TM es transcription mark; y MD, discourse marker.	233
7.4. Tabla de especificación de los tipos contextuales	240
7.5. Tabla de especificación de los tipos contextuales que presentan buenos resultados en la desambiguación	242
7.6. Tabla de reglas contextuales que han dado un resultado negativo	243
7.7. Tabla de especificación de los tipos contextuales con resultado ambiguo	244
7.8. Tabla de frecuencia marcador vs. categoría	246
7.8. Tabla de frecuencia marcador vs. categoría	247
7.8. Tabla de frecuencia marcador vs. categoría	248
7.9. Tabla de aciertos de marcador sobre categoría	249
7.10. Tabla de fallos de marcador sobre categoría	250
7.11. Tabla de frecuencia ambigua de marcador sobre categoría	252
8.1. Distribución de la parte informal de C-Oral-Rom	260
8.3. Distribución de la parte formal de C-Oral-Rom	260
8.4. Tabla de frecuencias de MMDD de Relaciones Discursivas	266
8.4. Tabla de frecuencias de MMDD de Relaciones Discursivas	267
8.4. Tabla de frecuencias de MMDD de Relaciones Discursivas	268

8.4.	Tabla de frecuencias de MMDD de Relaciones Discursivas	269
8.4.	Tabla de frecuencias de MMDD de Relaciones Discursivas	270
8.4.	Tabla de frecuencias de MMDD de Relaciones Discursivas	271
8.4.	Tabla de frecuencias de MMDD de Relaciones Discursivas	272
8.4.	Tabla de frecuencias de MMDD de Relaciones Discursivas	273
8.5.	Tabla de frecuencias de MMDD de Evaluación Emocional	273
8.5.	Tabla de frecuencias de MMDD de Evaluación Emocional	274
8.5.	Tabla de frecuencias de MMDD de Evaluación Emocional	275
8.5.	Tabla de frecuencias de MMDD de Evaluación Emocional	276
8.5.	Tabla de frecuencias de MMDD de Evaluación Emocional	277
8.6.	Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción	277
8.6.	Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción	278
8.6.	Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción	279
8.6.	Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción	280
8.6.	Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción	281
8.6.	Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción	282
8.6.	Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción	283
8.6.	Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción	284
8.7.	Tabla de frecuencias de MMDD de Evidencialidad	285
8.8.	Tabla de frecuencias de MMDD con significado ilocutivo	285
8.8.	Tabla de frecuencias de MMDD con significado ilocutivo	286
8.8.	Tabla de frecuencias de MMDD con significado ilocutivo	287
8.9.	Tabla de frecuencias de MMDD de Deixis	288
8.9.	Tabla de frecuencias de MMDD de Deixis	289
8.9.	Tabla de frecuencias de MMDD de Deixis	290
8.9.	Tabla de frecuencias de MMDD de Deixis	291
8.9.	Tabla de frecuencias de MMDD de Deixis	292

Introducción

Parece fuera de toda duda que el funcionamiento del lenguaje dentro de la mente es uno de los retos científicos más importantes del nuevo siglo; no solo por lo que este conocimiento pueda esclarecer sobre el gran misterio del funcionamiento del cerebro, sino por la necesidad que la Inteligencia Artificial tiene de un marco teórico y metodológico sobre el lenguaje desde el cual hacer sus programas e integrarlos dentro de una arquitectura general. No obstante, la contestación a la pregunta de si se podrá algún día formalizar el lenguaje natural con el lenguaje de la lógica o de las matemáticas todavía está fuera hoy por hoy de nuestro alcance. La relación entre los lenguajes formales y las lenguas naturales, o mejor dicho, la implementación de las últimas en los primeros es todavía un hecho que, de momento, queda reservado al mundo de la ciencia ficción. No deja de ser curioso y hasta paradójico que muchos físicos, entre ellos Richard Hawkins en su libro *Teorías del todo*, se hayan sorprendido de que las leyes del universo se puedan expresar con unas matemáticas bien sencillas (a la vista están las leyes de Newton y la fórmula de Einstein), y que sin embargo, las lenguas naturales se resistan a las puertas del siglo XXI a ser formalizadas en términos matemáticos. Las matemáticas y el lenguaje parecen ser códigos reservados a codificar dos conocimientos diferentes dentro de la mente: el mundo objetivo y el mundo subjetivo respectivamente. La expresión de uno en términos de otros y sus relaciones de equivalencia así como su interacción dentro del cerebro es otro misterio más sin resolver hoy en día para la ciencia. En la segunda mitad del siglo XX, con la implementación de las gramáticas formales chomskyanas, se originaron en el seno de las Ciencias Cognitivas unas expectativas demasiados optimistas que poco a poco se han ido reformulando en posturas más prudentes y modelos de lenguaje aplicados a dominios restringidos. Si tuviéramos que localizar con la precisión de un cirujano a punto de operar dónde se encuentra la causa de que todavía las máquinas no hablen, parece claro que en el centro de la diana se situaría el gran problema de la interpretación del significado. Y cuando hablamos del significado, no nos referimos solo al de las palabras, sino también al de estas mismas cuando se combinan en oraciones conformando textos o discursos.

Por lo tanto, una vez que el nivel sintáctico, más o menos, deja de ser un reto para la formalización, ahora debemos centrarnos en la aprehensión del significado y su relación con la forma lingüística. Un significado que, a medida que más se profundiza en él, más asoma su faceta cognitivo-simbólica, sociocultural y psicológica-individual. Estamos acostumbrados a escuchar que el significado de las palabras es convencional, y que por eso existe el lenguaje, porque nos hemos puesto de acuerdo en que, cada vez que la grafía o la cadena de sonidos *mesa* llegue a nuestros oídos, nosotros entendamos todos por igual lo mismo. Con el valor de verdad de este último enunciado, los ingenieros en informática y telecomunicaciones abordan el problema del significado de las lenguas y su codificación por la forma lingüística como si de algo claro, objetivo y preciso se tratara. No obstante, como hemos dicho, las lenguas naturales, en oposición a los lenguajes matemáticos, son sistemas de simbolización de significados tan flexibles como ambiguos a partir de su forma lingüística. Las normas culturales y sociales nos guían en la interpretación de las palabras y de los enunciados, sin embargo, tarde o temprano aflora esa dimensión del lenguaje que se opone a ser formalizado, predecible y sin la cual no existiría la poesía.

Así pues, hoy por hoy, el gran problema al que se enfrentan tanto la Lingüística Computacional como la Inteligencia Artificial a la hora de plantearse el procesamiento del lenguaje natural es el de la interpretación del significado, y más concretamente, la interpretación de los significados de las oraciones, o mejor dicho, de los enunciados, los cuales no son otra cosa que las oraciones pronunciadas en unas coordenadas espacio-temporales, en un contexto determinado. El estudio de los enunciados, su significado y la relación entre ellos en el discurso constituyen el centro de atención de la mayor parte de los estudios adscritos a la Pragmática Teórica. Dicha corriente se gesta en sus orígenes en gran medida por oposición a la Lingüística Formal, Generativa o Chomskyana, y es precisamente la formalización el aspecto que menos les preocupa a estos estudios de corte radicalmente funcionalista (?). La consecuencia de ello es la dificultad de aplicar los conocimientos procedentes de esta disciplina a otras parcelas de conocimientos de naturaleza aplicada como las que acabamos de señalar, la Enseñanza de Español a Extranjeros o la Psicolingüística.

La formalización es una de las razones por las que, como decimos, la Lingüística Computacional todavía está lejos de incorporar todo el conocimiento generado en esta disciplina a sus modelos de lenguaje. Solo de forma paulatina la Pragmática Computacional está empezando a implementar algunos marcos teóricos procedentes de la Pragmática Teórica, como por ejemplo la Teoría de los Actos de Habla o la localización de marcadores discursivos en los textos de manera automática; sin embargo, todavía existen muchos fenómenos que tienen lugar en el nivel enunciativo de la lengua que la Pragmática Computacional no

ha contemplado, como por ejemplo, la modalización discursiva, y que ejercen una gran influencia en la interpretación de los valores de verdad de los enunciados.

Al mismo tiempo, el hecho de que la Pragmática carezca de un marco teórico unificado hace que todavía no existan, en la anotación de corpus, estándares sobre etiquetado pragmático en los que se puedan explicitar conceptos y fenómenos que son claves en esta dimensión de la lengua. Buena parte de las anotaciones de corpus con información pragmática que se han hecho han tenido lugar en el terreno de la Inteligencia Artificial, en concreto en los sistemas de diálogo hombre-máquina, y sus esquemas de etiquetado no trasluce una postura lingüística detrás, sino que son implementaciones vagas de conceptos desligados entre sí y absolutamente adaptados al dominio en que trabajan. Esta manera de trabajar no nos debería extrañar si ni siquiera hay un acuerdo en la comunidad científica sobre qué se considera anotación pragmática. Estas deficiencias se agudizan más en el panorama español, ya que hoy por hoy, no se ha contemplado por parte de las instituciones que poseen grandes corpus, y entre ellos grandes corpus de español oral, marcar al menos como categoría los marcadores del discurso, y de paso, plantearse los problemas derivados de hacer esta tarea lo más automática posible.

A la vista de estos datos, es de rigor hacer alguna reflexión que nos obligue a comprender algunas de las razones por las que los informáticos que trabajan en procesamiento del lenguaje natural y que no ignoran estos conocimientos procedentes de la Pragmática (los cuales conforman un grupo muy reducido) ofrecen casi tanta resistencia como los propios lingüistas a la hora de fijar en una etiqueta el significado de un marcador del discurso o del acto de habla de un enunciado. Ya que en estos casos, dicho significado no es tan automático, veloz e inconsciente como en el caso del léxico, sino que es producto de la interacción con normas socio-culturales y, en último término, de nuestra propia psicología. Así por ejemplo, ¿cómo debo interpretar/etiquetar el enunciado *Habría que bajar la basura* dentro del contexto de la convivencia conyugal? ¿como una orden o simplemente como una aseveración?

Es por ello que, dentro de la enumeración de los retos o problemas de la Pragmática, deben tratarse como cuestiones claves la predictibilidad en la interpretación del significado, el concepto de contexto, y la codificación lingüística en términos de relevancia durante el proceso de comunicación.

En efecto, la Pragmática, en tanto que estudia el uso del lenguaje en su contexto, es una disciplina sujeta a la interpretación. La Filosofía de la Ciencia la clasificaría como un campo de investigación perteneciente a las disciplinas propias del terreno de la subjetividad (ARTIGAS, 1999). Buena parte de la Pragmática se ocupa del conocimiento implícito en los mensajes lingüísticos, esto es, aquello que se infiere pero que no está explicitado ver-

balmente. El simple hecho de plantearse esta parte del significado como objeto de estudio supone ya en sí mismo un problema para un paradigma científico que sienta sus bases en la observación empírica y en la predictibilidad. El significado final de una emisión lingüística para un receptor es el resultado de una inferencia mental, no se puede observar directamente y muy pocos dirían que es predecible, ni siquiera en situaciones o contextos sometidos a un alto nivel de descripción.

Para complicar aún más el acercamiento *empírico* a esta materia, debemos añadir que el material de trabajo con el que se llevan a cabo las investigaciones en Pragmática está conformado por corpus. En estos conjuntos de textos, el contexto ha desaparecido por completo y solo permanecen algunas de sus huellas reflejadas en la información de las cabeceras y en algunas palabras como, por ejemplo, las expresiones deícticas, entre otras formas lingüísticas. Algunas corrientes relativistas defienden que es imposible predecir en estos momentos el significado, por ejemplo, de un acto de habla indirecto, simplemente contando con la información lingüística de un enunciado. Como sabemos, es el contexto el que ayuda a que algunos significados y no otros se activen durante el proceso de inferencia que tiene lugar durante la comunicación. Al mismo tiempo, la Lingüística Interactiva (MONDADA, 2001) también apunta en esta dirección al señalar que los significados se construyen y se negocian de forma activa por parte de los participantes durante el proceso de interacción, como ya hemos señalado anteriormente.

Así pues, el desafío más importante para las disciplinas que se plantean la comprensión y generación de lenguaje natural está centrado en la sistematización o formalización de la relación forma lingüística, significado y contexto. Este último concepto está sujeto a diferentes parcelaciones por parte de los diversos enfoques teóricos que se han ocupado de él. Según algunas corrientes de la Psicología Cognitiva (GARCÍA, 1991) y de la Inteligencia Artificial (JOHNSON-LAIRD, 1993), los contextos serían marcos mentales que ayudan al individuo a interpretar la información del medio, como por ejemplo, una conversación, en términos de valores de verdad, pero ¿cómo dar cuenta de esta realidad mental entonces?

Por último, y para terminar la enumeración de los retos más importantes que se plantea la Pragmática Computacional, debemos tener en cuenta que la comunicación humana es un proceso inferencial y que, tal y como han puesto de manifiesto los investigadores Sperber y Wilson (SPERBER y WILSON, 1994) a través de la formulación de la Teoría de la Relevancia, solo codificamos la parte más relevante de lo que queremos decir. Si tenemos en cuenta la propuesta de estos autores, es de obligada reflexión decidir también qué parte del significado es posible recuperar realmente.

Estas cuestiones que someramente acabamos de apuntar sobre los grandes problemas que conlleva la aproximación computacional al significado es objeto de un candente debate

en la investigación teórica pero, por contrapartida, apenas es un tema considerado no solo en las empresas de ingeniería lingüística sino en la investigación pública dedicada al procesamiento de las lenguas, la cual, a día de hoy, está dominada por investigadores formados en ingeniería informática que prefieren aproximaciones al procesamiento del lenguaje natural desde modelos probabilísticos, y donde el lingüista, lejos de tener un papel activo en la planificación de estrategias de programación, cumple el papel de corregir o depurar los programas hechos por los ingenieros que no gozan precisamente de una elevada formación lingüística, ni en su plano teórico ni en su plano más descriptivo.

Por todas estas razones, la primera motivación de este trabajo ha sido la de acercar a las disciplinas que tratan con el lenguaje el conocimiento que se está construyendo en torno a la dimensión pragmática de la lengua; nos referimos, más concretamente, a la Lingüística Computacional, la Enseñanza de Español a Extranjeros, la Sociolingüística de corte cuantitativo y los sistemas de diálogo hombre-máquina.

La vía de trabajo propuesta en esta tesis ha tenido en cuenta todas las limitaciones que se acaban de exponer en torno al significado inferido o no explicitado y defiende, como punto de partida, la consideración de la dimensión pragmática en el procesamiento del lenguaje natural a través de la anotación pragmática de corpus.

Desde un punto de vista más concreto, el objetivo de esta investigación ha sido el de introducir información de tipo pragmático en el corpus oral del español C-ORAL-ROM. La información que se refleja en la etiqueta de las unidades lingüísticas que vamos a marcar remite a fenómenos que se podrían situar en el interfaz semántico-pragmático de la lengua; dichos fenómenos tienen una naturaleza cognitivo-social y afloran en el momento de la enunciación. En concreto, nos interesa saber cómo se marcan lingüísticamente en una lengua las emociones, el razonamiento, la modelización discursiva de la realidad, las fuentes en la que apoyamos el valor de verdad de nuestros enunciados, la alusión al oyente y al discurso, la acción y, por último, la convencionalización de la lengua. Dichos fenómenos se han denominado, respectivamente, desde la Lingüística pero no solo: lenguaje emocional, relaciones discursivas, modalización discursiva (atenuación e intensificación), evidencialidad, deixis social y discursiva, actos de habla y unidades fraseológicas.

Los objetos de estudio ya en el nivel más puramente gramatical que hemos elegido para proyectar el modelo de anotación han sido los constituyentes extraoracionales, denominados tanto por la gramática tradicional de corte estructuralista, como por las vertientes formalistas del lenguaje, como por el neoestructuralismo: adverbios oracionales, interjecciones, marcadores discursivos, vocativos, sintagmas preposicionales externos y adjuntos. En nuestro modelo de anotación, estas formas léxicas o lexicalizadas serán clasificadas semánticamente a la luz de los fenómenos nombrados y etiquetadas en el corpus cuan-

do aparezca alguna de sus ocurrencias. Para los no especialistas en Pragmática, tenemos que decir que el modelo lingüístico que ofrecemos aquí recoge fenómenos que los investigadores en esta disciplina están descubriendo y que no es un modelo cerrado compuesto por unidades discretas que guarden relaciones de oposición entre sí, como por ejemplo los sistemas descritos en el nivel fonológico.

En el futuro, este esquema de anotación se aplicará a nivel intraoracional, ya que los fenómenos que aquí son objeto de análisis se manifiestan en la lengua no solo a través de los constituyentes periféricos de la oración sino también a través de otros niveles de la gramática como la sintaxis, la prosodia o la morfología. Por ejemplo, la atenuación se muestra en el nivel morfológico en el caso de *¿Tomamos una **cervecita**?*. La introducción de este tipo de anotación en los corpus puede ayudar a la consideración de estos fenómenos por parte de la Inteligencia Artificial y de la Lingüística Computacional, fenómenos que, como decimos, pueden estar codificados lingüísticamente, y que aluden a la naturaleza subjetiva de nuestro conocimiento de la realidad.

Hemos estructurado este trabajo en cinco partes denominadas de la siguiente manera: **de la teoría, de la descripción, del procesamiento, de la aplicación y de la interpretación**. Veamos cuáles son las líneas de trabajo que se tratan en cada una de estas partes.

La primera parte, **De la Teoría**, está conformada por tres capítulos. En el primer capítulo hablamos de la relación entre cognición, conducta social y uso del lenguaje en el ser humano. En el segundo capítulo, tratamos las aportaciones teóricas más relevantes que han tenido lugar dentro del terreno de la Pragmática. En concreto hablaremos de las siguientes corrientes: la Teoría de la Relevancia, el modelo de cooperación comunicativa de Paul Grice, el modelo de cortesía propuesto por Levinson y Brown, la Teoría de los Actos de Habla y, por último, la Teoría de la Argumentación. Finalmente, el tercer y último capítulo de este módulo teórico está dedicado a los marcadores discursivos, objeto de estudio por excelencia en los estudios pragmáticos; en él trataremos cuestiones relativas a su definición, significados, propiedades gramaticales, clasificación semántica y tratamiento de los mismos dentro de la Lingüística Computacional.

La segunda parte de este trabajo está conformada por lo que hemos denominado el **módulo descriptivo**, el cual ocupa dos capítulos: en el primero de ellos, se presenta el modelo de anotación pragmática que hemos diseñado, Pragmatext. En él se describen los fenómenos de naturaleza semántico-pragmática que son objeto de nuestro análisis; en el segundo capítulo, en cambio, hablaremos de los problemas de formalización del significado de los marcadores discursivos del corpus C-ORAL-ROM, y ofrecemos una clasificación semántico-pragmática de los mismos a la luz de los fenómenos descritos.

La tercera parte de este trabajo, **el módulo computacional**, está compuesto por dos

capítulos. En el primero de ellos, nos ocupamos de la presentación de las características más relevantes del corpus oral del español C-Oral-Rom y de su etiquetado en formato XML; el segundo capítulo está dedicado a exponer las estrategias adoptadas para resolver el problema de la identificación, desambiguación y etiquetado semiautomático de estas partículas. En él presentaremos **MiDaStagger**, la herramienta de anotación que hemos elaborado para etiquetar los marcadores discursivos del corpus.

En la cuarta parte, **De la aplicación**, hemos querido demostrar la utilidad de los corpus anotados con este tipo de información, y para ello mostramos dos posibles aplicaciones a las siguientes parcelas de conocimiento: los estudios de Sociolingüística de carácter cuantitativo y la Enseñanza de Español para Extranjeros. Para ello presentamos nuestro interfaz Web **Inter Alia**, una aplicación informática de consulta de corpus para recuperar marcadores discursivos, ejemplos del corpus y frecuencias en función del género discursivo. Una herramienta de gran utilidad tanto para los lingüistas como para los alumnos no nativos que deseen aprender lenguas de una manera más individual y a través de la consulta de corpus.

Y finalmente, el último módulo, titulado **De la interpretación**, está conformado por el capítulo diez y las conclusiones. En estos dos apartados reflexionamos sobre los retos que para la Inteligencia Artificial supone la integración de los conocimientos que se están generando desde la Pragmática Teórica, fundamentalmente en lo que a marcadores del discurso se refiere.

Antes de que el lector comience la lectura de este trabajo, nos gustaría hacer una pequeña aclaración. Este es un trabajo interdisciplinar, con lo cual, imaginarse un lector ideal durante la redacción de esta tesis ha sido una tarea ardua, ya que como sabemos lo que es obvio para unos es absolutamente lejano e ignorado por otros. En principio, o al menos este sería nuestro deseo, muchas personas de diferentes formaciones académicas pueden interesarse por este trabajo; por ello, hemos pretendido, dentro de la brevedad, hacer una breve presentación de todas las parcelas de conocimiento que se cruzan en este trabajo. Este estudio propone una vía de transferencia de conocimientos de una parcela teórica a otras de naturaleza aplicada, con lo que el lector no encontrará aquí el extenso estado de la cuestión y el intenso debate teórico existente sobre cada una de las cuestiones que aquí se tratan, sobre todo las concernientes a la Pragmática Teórica y en concreto, en torno a la relación entre Semántica y Pragmática. Si el lector es ducho en un fenómeno en concreto, como por ejemplo en la modalización discursiva, aquí accederá a nuestra propia visión de la misma, después de una breve introducción; si por el contrario no lo es, y siente

inquietud por profundizar en la cuestión desde un punto de vista teórico, cortésmente le remitimos a la bibliografía citada en cada capítulo.

Parte I

De la teoría

Capítulo 1

Cognición y comunicación, un marco para la dimensión pragmática de la lengua

Una palabra perdida es, tal vez, un acceso a la realidad perdido.

Teoría de la Inteligencia Creadora. J. A. Marina.

1.1. Introducción

Durante la introducción, hemos definido la Pragmática como el campo del saber que estudia el modo en que los seres humanos utilizan el lenguaje en los diferentes contextos de su vida diaria. Es en esta dimensión de la lengua donde más se evidencian las relaciones que el lenguaje guarda con otras dimensiones de la realidad como nuestra manera de ver y sentir el mundo o nuestra forma de relacionarnos con otros seres humanos. Dicha relación es la causa de que algunas propuestas que forman parte imprescindible de la fundamentación teórica de la Pragmática terminen trascendiendo el análisis de la forma lingüística e incluyan o bien el plano de lo sociocultural o bien el plano lo cognitivo en su concepción final del lenguaje y de la comunicación. A pesar de que existen iniciativas que pretenden unir estas dos dimensiones (la cognitiva y la sociocultural) dentro de un modelo general (ESCANDELL, 2004), esta dicotomía en la investigación pragmática solo es una manifestación más, entre otras, de un intenso debate científico en torno al universalismo o relativismo cultural. La Ciencia Cognitiva, hábida de buscar generalizaciones estables a la especie, dejó de lado en su primera etapa el nivel sociocultural. En este sentido, las corrientes más formalistas del lenguaje han dejado fuera también el problema del significado contextual. No obstante, dentro de lo que podríamos denominar el cognitivismo de segunda generación, también existen líneas de investigación que pretenden tender un puente entre lo

cognitivo, lo social y lo lingüístico. Como ejemplo de la relación entre cognición y lenguaje no podemos dejar de señalar la Lingüística Cognitiva. Esta disciplina intenta dar cuenta de cómo interactúan la mente, el cuerpo y el lenguaje, y “busca activamente las correspondencias entre el pensamiento conceptual, las experiencia corpórea y la estructura lingüística, al considerar que las categorías lingüísticas no son autónomas respecto de la organización conceptual y general y de los mecanismos de procesamiento mentales”(CUENCA y HILFERTI, 1999) . Al mismo tiempo, dentro de la relación entre lo social y lo lingüístico, hablaremos de terrenos de conocimiento al margen de las Ciencias Cognitivas, tales como la **Microsociología**, la **Etnometodología** y el **Análisis de la Conversación**, que han dado también sus frutos a la hora de estudiar la relación entre lo social y los usos del lenguaje. No obstante, a pesar de estos avances, la Lingüística Computacional solo de forma reciente se está preocupando por la formalización del significado contextual.

Hace cinco años, en el congreso LREC2004, un informático de formación matemática me explicó que cuando él trataba con el lenguaje no veía palabras, sino que veía cadena de caracteres. Me gustaría con este capítulo, mostrar las raíces cognitivas y sociales del lenguaje; explicar, en definitiva, cómo el lenguaje, lejos de ser simples cadenas de caracteres, codifica conocimiento de la realidad, de la interacción social, del mismo intercambio comunicativo, y es además, un instrumento cognitivo para hacer comprensibles unas realidades en términos de otras.

La línea argumentativa que vertebra este capítulo está orientada a ilustrar el modo en que interactúan entre sí estos **tres niveles: lo lingüístico, lo socio-cultural y lo cognitivo-mental**. Los fenómenos que queremos etiquetar son un buen ejemplo de la interacción que existe entre cerebro-mente, sociedad-cultura y lenguaje, y así lo ilustraremos a lo largo de nuestra exposición. En este capítulo, presentaremos las raíces cognitivo-sociales del ya conocido tema de la subjetividad en la lengua, e introduciremos brevemente la dimensión cognitivo-social de los fenómenos que conforman nuestro modelo de anotación pragmática y que posteriormente, en el capítulo cuatro, describiremos con más detalle desde una perspectiva lingüística. Con un afán didáctico, vuelvo a enumerar los fenómenos que se tratan en este trabajo: **evidencialidad, metáfora, certeza, acción, emociones, razonamiento y deixis**. La categoría pragmática en la que vamos a centrarnos para el estudio de dichos fenómenos son los **marcadores del discurso**. Aunque ya lo hemos señalado, volvemos a reiterar que la Lingüística Computacional, la Inteligencia Artificial y la Enseñanza de Español a Extranjeros deberán conocer el origen y trascendencia de estos fenómenos si quieren plantearse el problema del significado de los enunciados y de los textos. El compo-

nente común de estas tres disciplinas es la reflexión sobre el lenguaje desde la perspectiva del desconocimiento del código, aunque con la diferencia de que la máquina y el hablante no nativo guardan recursos de diferente naturaleza para enfrentarse a una nueva codificación lingüística procedente de una nueva interpretación de la realidad.

1.2. La construcción de la subjetividad: el observador de metáforas

No deberíamos olvidar que lo que observamos no es la naturaleza misma, sino la naturaleza determinada por la índole de nuestras preguntas. W. Heisenberg, 1959 en (MARINA, 1993)

La Lingüística se ocupa fundamentalmente de la relación entre la forma lingüística y el significado, y de los problemas derivados de esta relación. La realidad deja de ser objeto de estudio para el lingüista en el momento en que ha sido convencionalizada en el significado. Debemos al estructuralismo sausseriano una conceptualización estable de la relación entre el referente y el signo lingüístico (compuesto, como sabemos, de significante y significado). Solo en las últimas décadas del siglo XX, la Teoría de la Enunciación ha hecho hincapié en las marcas de subjetividad presentes en el lenguaje, como por ejemplo, la deixis o la actitud del hablante respecto de lo enunciado (TORDESILLAS y NEGRONI, 2001). Así, por ejemplo, es un hecho bastante usual llamar al telefonillo de casa y contestar a la pregunta de *¿quién es?* con el pronombre de primera persona: *yo*. Asimismo, es bastante común acompañar el valor de verdad de nuestros enunciados con un valor emocional; con lo que podemos decir: ***Por fin***, *un presidente de los EEUU no es rico*, o ***Afortunadamente***, *los políticos ya no podrán llevarse el dinero público robado a los paraísos fiscales*.

No obstante, aunque esta visión de la lengua ha permitido adentrarnos en el cambiante mundo de la referencialidad de los signos lingüísticos, estas corrientes enfocan el problema de la subjetividad solamente desde un punto de vista lingüístico, cuando en realidad este fenómeno tiene una base cognitivo social que pasaremos a explicar aquí.

Un enunciado codifica información sobre las acciones, sobre los valores de verdad, sobre el grado de certeza, sobre las fuentes de conocimiento, sobre la relación lógica que guarda con los enunciados que le rodean, y sobre las emociones que dicho enunciado ins-

pira. Esta es pues la información que contribuye al significado cognitivo-sociocultural de los enunciados y de las palabras. Y esta es pues también la información que debe tenerse en cuenta en los modelos de lenguaje en la IA, no solo el valor de verdad del enunciado. Pero antes de explicar esto más detenidamente, me gustaría reflexionar primero sobre la percepción del realidad por parte del ser humano y su codificación por los sentidos. Langacker, uno de los padres de la Lingüística Cognitiva, ha insistido bastante en esta cuestión: “The world is not something objectively given, it’s something **construed** by the human cognition.” (CUENCA y HILFERTI, 1999)

Dentro de la construcción de la subjetividad, como sabemos, los sentidos son el primer sesgo ineludible. Centrémonos en el sentido de la vista, por ser este el que, en el caso de los videntes, contribuye a la construcción de lo que el ser humano entiende por realidad.

Los estudiosos del universo afirman que con la luz captada por el ojo humano solo podemos percibir el cuatro por ciento de la realidad y señalan que estamos ciegos a los colores del universo (MURTAGH, 2004), unos colores que solo ahora comenzamos a contemplar con la difusión de las imágenes por parte de la NASA de grandes telescopios como el *Hubble*. La Ciencia Cognitiva también se ha centrado mucho en el estudio del ojo (PINKER, 1997), y las investigaciones realizadas sobre la percepción del color en diferentes culturas son una buena muestra de ello (GADNER, 1985). En principio, podríamos pensar que existe una parte de la realidad a la que accedemos por los sentidos y otra a la que no, y que el lenguaje se ocuparía de crear relaciones estables, como decíamos, entre los signos lingüísticos y entre lo conocido y en todo caso lo imaginable por nuestra mente (vida extraterrestre, espíritus, Dios, etc.). Sin embargo, la Física Cuántica, en sus incursiones al mundo de la divulgación del conocimiento, nos ha explicado que la realidad depende del observador, o mejor dicho, la realidad es el fruto del acto de observación por parte del observador (CHOWN, 2006). En palabras de Antonio Marina: “No hay información sin receptor. Sin el ojo no existe el color. [...] El hombre posee mentalmente la realidad” (MARINA, 1993).

El órgano visual del ser humano ya es en sí mismo una interpretación de lo real, bien distinta de la interpretación de la realidad que haríamos si tuviéramos un sistema visual bifocal, como el camaleón, o de si viéramos/percibiéramos el mundo a escala atómica. Las cosas que existen existen porque nosotros las vemos así, o mejor dicho, solo existe la información que los sentidos son capaces de captar y de ser codificada posteriormente por el sistema nervioso dentro del cerebro. Los sentidos son nuestra primera fuente de conocimiento y el primer constructor de subjetividad en el ser humano. Según el modelo dialéctico contextual en psicología cognitiva, el conocimiento tiene su origen en la interacción entre el sujeto y el medio (GARCÍA, 1991). Sabemos que la experiencia determina de forma importante la manera en que vemos las cosas. Los humanos que viven en las

ciudades de tipo occidental tienen una experiencia de las líneas horizontales y verticales diferente de la de los habitantes de la jungla, de quienes se dice que no experimentan las mismas ilusiones espaciales (YOUNG, 1976).

El lenguaje codifica la visión de la realidad codificada por los sentidos. Como por ejemplo, nuestra concepción del tiempo. El tiempo desde un punto de vista de la física heredera de la teoría tradicional de la relatividad, tiene una dimensión relativa. Las partículas que se encuentran en el universo se mueven a diferentes velocidades. Nuestra mente, que por encima todo es un receptor del tiempo, captura solamente las partículas que se mueven a una velocidad determinada, y como resultado nos devuelve una concepción lineal del tiempo, que en nada tiene que ver con la concepción del tiempo que está descubriendo la Física Cuántica al estudiar el movimiento de partículas que se mueven a velocidades cercanas a la luz. Las lenguas codifican nuestra concepción lineal del tiempo: presente, pasado y futuro. Como el tiempo es un concepto abstracto, suele expresarse muchas veces en términos del espacio. Esta es la función a nivel cognitivo de la metáfora en las lenguas. En palabras de Cuenca (CUENCA y HILFERTI, 1999) “la metáfora es un mecanismo cognitivo que se utiliza para procesar información abstracta a partir de conceptos más simples y familiares”, que normalmente coinciden con los del dominio de lo concreto, aquello que se puede tocar, medir, oler, ver y oír. Es por ello que es difícil saber si el marcador discursivo *anteriormente* en el enunciado *Anteriormente, hablé de la relación entre política y acumulación del capital* se refiera tanto a un punto en el espacio, como a un punto en el tiempo. Los ordenadores del discurso, como por ejemplo, las palabras *primero*, *segundo* y *tercero*, también están transmitiendo una concepción lineal del tiempo, que en los textos se materializa con puntos en el espacio diferentes en la presentación de argumentos.

Así pues, el lenguaje, además de ser un sistema de símbolos, también es un instrumento en sí mismo para conocer la realidad. Una realidad que siempre está por descubrir, ya que nuestro conocimiento de la misma siempre va a ser limitado. A través de la metáfora se contribuye al proceso de designación lingüística, a la organización de nuestra experiencia y, en último término, a la construcción de nuestra realidad. Tal y como se señala en (CACCIARI, 1995): “There is a parallel between our propensity to perceive things in terms of other things and our propensity to expand the meaning of words and expressions by figurative speech”. En este sentido, no podemos dejar de señalar estudios como el de *La metáfora como recurso epistémico* (ZAVADIVKER, 2005), en los que se nos explica cómo las metáforas del lenguaje científico también determinan ya desde su concepción una visión del fenómeno analizado. Uno de los ejemplos más famosos, como sabemos, es la concepción del cerebro como un ordenador químico. También es justo señalar que no siempre podemos hacer comprensibles los nuevos descubrimientos a partir de nuestro conocimien-

to cotidiano de la realidad. De hecho, muchos físicos se quejan de que ciertos conceptos de la nueva física no puedan transmitirse didácticamente a falta de una metáfora que los haga comprensibles. La mayoría de las veces el mismo proceso de designación de los nuevos descubrimientos o conceptos es metafórico; pensemos, por citar uno, en la construcción lingüística *un agujero negro*.

Con estos ejemplos procedentes del mundo científico, esperamos haber ilustrado suficientemente el origen metafórico del lenguaje. Esta perspectiva no hubiera sido posible si no fuera por la investigación realizada por la Lingüística Cognitiva, y en concreto, por sus investigadores inaugurales, George Lakoff and Mark Johnson (FAJARDO, 2006) (LAKOFF y JOHNSON, 1980).

Pasemos ahora a tratar otros planos del lenguaje, como su relación con la acción, la lógica clásica y las emociones a la hora no solo de transmitir sino también de intentar construir nuestra subjetividad.

1.2.1. Pensamiento, lenguaje y acción: acciones codificadas y acciones inferidas

La información sobre la realidad representada en la mente se manipula con el fin de hacer una predicción con miras a la acción. En palabras de Jose Antonio Marina (MARINA, 1993): “No actuamos para conocer sino que conocemos para actuar [...] Lo que define una acción es la intención, el proyecto y plan que los guía”.

El filósofo del lenguaje Jean Austin centró su obra en poner de manifiesto que los enunciados no solo codifican acciones *Juan come patatas*, sino que también realizan acciones, como por ejemplo, el enunciado, *Sí, quiero* dicho en una boda o *Declaro la guerra a Alemania* dicho por el presidente de otro país. Tanto es así que, en la comprensión total de un enunciado, buena parte de su interpretación pasa por inferir cuál es su fuerza ilocutiva, esto es, cuál es la acción que está codificando de una manera indirecta (una orden, una petición, un ruego, una disculpa, etc.). Inferir el significado de un enunciado significa hacer una predicción sobre la intención con la que el hablante ha dicho lo que ha dicho. La Pragmática Computacional debe incluir el significado ilocutivo de los enunciados como una parte relevante del significado de los enunciados. En este trabajo, haremos una primera aproximación a este tema ya que algunos de los marcadores del discurso que vamos a etiquetar codifican alusiones a las acciones contextuales, como por ejemplo, *ánimo* o *que en paz descanse*.

1.2.2. Verdad, certeza y evidencialidad.

Como sabemos, los conocimientos albergados en nuestra mente sobre el mundo distan mucho de ser científicos (RÁBADE, 1995). En realidad, en lugar de verdades deberíamos hablar de creencias. Además, a medida que los conocimientos son más abstractos, estas creencias difieren más de un individuo a otro. Fijémonos en que sólo cuando creemos algo pasa a ser verdad para nosotros. Debemos acostumbrarnos a pensar que los valores de verdad de los enunciados son contextuales, como lo son la verdad obtenida de los razonamientos lógicos que hagamos con ellos. El enunciado *Dios existe* será verdadero o falso dependiendo del sistema de creencias de cada subjetividad; y solo así podemos procesar el significado de razonamientos como *Gracia a Dios, soy ateo*. A esto debemos añadir que el objeto conocido siempre es una proyección de nuestro propio conocimiento de la realidad, y este conocimiento nos ha sido transferido culturalmente. Tal y como señala Sergio Rábade (RÁBADE, 1995):

“La sociedad y la cultura es un universo simbólico que ofrece a disposición de los individuos un patrimonio de conocimiento. Las sociedades producen sus propias interpretaciones y representaciones de los eventos, sus propias teorías y explicaciones [...] Mediante el lenguaje, la sociedad se expresa, se comunica y transmite todo el bagaje de la cultura que con ese mismo lenguaje crea”.

Por lo tanto, si digo *Dios es amor*, este enunciado será verdad si se enmarca dentro del marco de determinadas religiones que unen estos dos conceptos en términos de equivalencia.

El hecho de que el cerebro genere una representación mental coherente de la realidad no significa que sea verdadera. Una de las diferencias más importantes entre los lenguajes formales y las lenguas naturales es el concepto de verdad. Nuestra concepción de la verdad se refiere a la relación de conformidad, correspondencia, coherencia, etc. que nuestro conocimiento guarda con lo conocido (RÁBADE, 1995). Como sabemos, Platón ideó la primera teoría de la verdad, teoría que constituye una parte imprescindible de la Semántica Formal y de la Lógica y mediante la cual los enunciados se dividen en verdaderos y falsos. Una proposición verdadera presenta los hechos tal y como son en realidad, mientras que una proposición falsa presenta cosas diferentes a los hechos (SERRANO, 1983). En un modelo computacional de lenguaje las oraciones son verdaderas o falsas, mientras que en las lenguas naturales, la relevancia se centra más bien en el grado de certeza que el hablante tiene con respecto a la verdad o no verdad de las proposiciones.

Este es un problema clave que la IA debe considerar a la hora de formalizar bases de

conocimiento, entendiendo por conocimiento una visión subjetiva de la realidad, parcial, pero coherente.

La Pragmática se ha centrado en los procesos de atenuación e intensificación del compromiso del hablante con la verdad de los enunciados. La Lógica también se ha ocupado de este fenómeno que ha denominado modalidad epistémica. Existen muchos marcadores del discurso que se encargan de intensificar o atenuar la certeza del hablante respecto de lo que está diciendo. Así por ejemplo, en el enunciado *El supuesto asesino de Marta del Castillo, Miguel, ha cambiado tres veces su declaración*, el adjetivo *supuesto* atenúa la certeza sobre el valor de verdad del enunciado. De la misma manera, en el enunciado *Sabemos, a ciencia cierta, que Marta del Castillo salió de su casa a las nueve de la noche*, el sintagma preposicional *a ciencia cierta* intensifica la certeza del hablante con respecto al valor de verdad del enunciado, y dicha atenuación o intensificación influye en la representación del conocimiento que tiene lugar en la mente del receptor de dicha información. Así es cómo se van conformando nuestras bases de conocimiento, del todo relativas por definición, ¿qué puede decir la IA respecto de este tema? Crear una inteligencia artificial en el sentido en que solo puede concebirse desde la literatura de ciencia ficción pasa por reconocer esta realidad, tan cotidiana para nosotros, como lo es el uso de un adverbio atenuante del valor de verdad como *supuestamente*.

Además, hay que añadir, que dentro de este mundo de conocimientos cambiante y relativo, tendemos a certificar la certeza para que sea creída por otros. Utilizar una fuente u otra de conocimiento confiere peso argumentativo, autoridad, o veracidad a la verdad de un enunciado. En las lenguas naturales, la fuente de conocimiento se codifica de diferentes maneras; algunas lenguas de Europa del Este incluso tienen un sistema de afijos verbales cerrado para expresarla, tal y como en su día nos explicó Jakobson (JAKOBSON, 1990 (1957)). Este fenómeno que codifica lingüísticamente la fuente de conocimiento se ha denominado **evidencialidad** y es un fenómeno universal a todas las lenguas. En nuestro sistema de anotación pragmática, Pragmatext, también nos ocupamos de marcar algunas formas lingüísticas que contienen significado evidencial. Así por ejemplo, con el marcador discursivo *por lo visto* en el enunciado *Por lo visto, Sarkozy ha dicho que Zapatero no es un muy inteligente*, el hablante marca lingüísticamente que la fuente de conocimientos en la que se apoya el valor de verdad de su enunciado no procede de su experiencia directa, o al menos así prefiere hacérselo creer a su interlocutor.

1.2.3. La lógica clásica y el sentido común: conectores lógicos y marcadores del discurso

Pasemos ahora a reflexionar sobre el funcionamiento de la mente a la hora de hacer predicciones. Como sabemos, la mente opera de forma inferencial haciendo razonamientos que le puedan conducir a una predicción representada en un valor de verdad. Por ejemplo, analicemos el razonamiento que le escuché en un programa de televisión a una persona que se dedicaba a limpiar todas las mañanas la playa de La Barceloneta *Si la gente no ensuciara la playa todas las noches, yo no tendría trabajo*. Una predicción es una inferencia, un razonamiento que conduce a una conclusión. Se suele decir que las predicciones deben hacerse sobre premisas verdaderas para asegurarse una conclusión o inferencia verdadera. Esto es lo que aprendimos de la Lógica Clásica. Pero parece que en el lenguaje cotidiano las cosas son bastante diferentes, la subjetividad está repleta de razonamientos que no necesariamente tienen que ser verdaderos de un modo absoluto, sí de un modo contextual. Así por ejemplo, en otro orden económico-social, es muy posible que este hombre sí encontrara otra clase de trabajos, invalidando el razonamiento hecho anteriormente.

La Lógica Clásica es un intento por abstraer el lenguaje a una lógica pura, formal, con la que supuestamente el cerebro operaría. Pero, hoy por hoy, los estudiosos del razonamiento mental deben resolver no solo la existencia de un lenguaje del pensamiento sino el del mejor instrumento para expresarlo, ya que hasta ahora el lenguaje es nuestra única puerta de acceso a este mundo desconocido, y hay bastantes discordancias entre la lógica formal y la lógica del sentido común, como acabamos de ver. La Inteligencia Artificial implementa modos de razonamiento procedentes de la Lógica Clásica ¿pero es así como razonamos los seres humanos en nuestra vida cotidiana?

A partir de los años setenta, buena parte de la IA y de la Psicología del razonamiento sintió una gran decepción por los modelos de razonamiento ofrecidos tanto por la Lógica Clásica. Estos investigadores alegaron que la Lógica no da cuenta de la manera de razonar en la vida cotidiana, que cuando la mente trabaja, lo hace siguiendo reglas que se basan en costumbres y convenciones sociales (LEGRENZI, 2000, (1ª Ed. 1998). Los contenidos influyen en nuestro modo de representar el problema y, como consecuencia, la abstracción de la estructura lógica no es suficiente para construir la epistemología natural. De hecho, algunos estudiosos incluso se han llegado a plantear si el cálculo lógico es un don natural. Legrenzi cita los estudios que Martín Brain realizó en 1978, de los que concluyó que la mayoría de las personas lleva a cabo con dificultad extrema la inferencia que los lógicos llaman *Modus Tollens*. Actualmente, la Psicología del razonamiento muestra una conciencia más general de la importancia del contexto social en la dirección de las decisiones y las acciones (LEGRENZI, 2000, (1ª Ed. 1998).

En lo que atañe a la Lingüística y más concretamente a la Pragmática, se han hecho intentos de relacionar conectores lógicos con conectores de lenguaje natural, sin buenos resultados (ESCANDELL, 1996). Así por ejemplo, la conjunción copulativa *y* no tiene los mismos significados en el lenguaje lógico que en el español. Por otra parte, además de sus posibles equivalentes lógicos, las lenguas ofrecen además una gama muy variada de formas lingüísticas con los que los enunciados se enlazan entre sí y que expresan operaciones de pensamiento. Según Blakemore, una marcador discursivo guía las instrucciones de interpretación del siguiente enunciado (MARTÍN y PORTOLÉS, 1999). Veamos un ejemplo que el profesor de pragmática de la Universidad Autónoma de Madrid, José Portolés, suele utilizar con frecuencia en sus clases para ilustrar este concepto. El enunciado *Es rico pero feo* puede conducir a la conclusión no explícita sino inferida: *No me casaré con él*. Sin embargo, si invertimos el orden de los enunciados pero conservamos el mismo conector: *Es feo pero rico*, la conclusión que se infiere es la contraria: *Me casaré con él*. Los enunciados despiertan inferencias en las mentes de los hablantes, y los marcadores del discurso cumplen el papel de restringir estas inferencias, de guiar los proceso de interpretación, y las conclusiones en el razonamiento mental. Además, hay marcadores que nos hablan de las operaciones que el cerebro hace a la hora de manipular los conocimientos. Los marcadores del discurso *generalmente* y *concretamente* nos hablan de los fenómenos de abstracción y concreción que realizamos a la hora de demostrar los valores de verdad de los enunciados.

Ejemplos de más marcadores que hablan de cómo vamos argumentando nuestra visión del mundo cuando hablamos son: *en consecuencia*, *por tanto*, *porque*, *ya que*, *si*, *dado que*, *visto que*, *como*, *en particular*, *de todos modos*, *etc.* Con estas palabras vamos razonando, creando conexiones lógicas entre enunciados que en principio no deberían tener una relación semántica si no es porque el hablante o la cultura quiere que así sea: *Si no votas, pues no te quejes*.

Ahora bien, ¿cómo representar desde un punto de vista lógico el significado de estas palabras? Diane Blakemore defiende la naturaleza computacional o procedural de estas palabras; sin embargo, paradójicamente, Anscombe y Ducrot, que también coinciden en la idea de que los marcadores restringen las inferencias del enunciado que viene a continuación, defienden que estas partículas no se pueden explicar con el lenguaje de la Lógica.

Vemos aquí, por tanto, una gran laguna en la representación lógica del lenguaje, con las implicaciones que conlleva a la hora de reflejar el razonamiento cotidiano, y el discurso entendido como una sucesión de valores de verdades contextuales.

1.2.4. La codificación lingüística de los estados emocionales

En el nivel fisiológico, la actividad mental tiene un correlato químico que determina buena parte de las emociones que sentimos. En palabras de Legrenzi: “Las pasiones del alma no son independientes de las maneras de funcionar de la mente. La felicidad y la infelicidad, el placer, el dolor, la alegría y el luto están estrechamente entremezclados en los procesos cognitivos que los generan”(LEGRENZI, 2000, (1ª Ed. 1998).

El estudio de las emociones ha recibido en las últimas décadas un gran impulso, no sólo por parte de la Psicología Cognitiva, sino también por parte de la Neurociencia y de la Inteligencia Artificial. Las emociones tienen también su correlato en el lenguaje. Localizar las formas léxicas que transmiten significado emocional sería una gran ayuda para los sistemas de comunicación hombre-máquina.

En la construcción del significado de los enunciados, la emoción que se les atribuya tiene un papel de gran importancia. Así pues, no nos interesa saber solo si un enunciado es verdadero o falso, o el grado de certeza que el hablante tiene con respecto al mismo, sino también es importante saber si, al hablante, los enunciados que produce o escucha le parecen bien o mal. Cuando aprendemos a hablar, las verdades también tienen un significado emocional negativo o positivo (*esto es bueno, esto es malo*) a través de las cuales se codifica el valor moral de las acciones de los enunciados y de las palabras. Los diez mandamientos son un buen ejemplo de ello. Ya hemos destacado algunos marcadores del discurso que dan cuenta del significado emocional de los enunciados. *Afortunadamente, desgraciadamente* o *felizmente*, colocados en posición extraoracional, son buenos ejemplos para ilustrar este fenómeno. En Pragmatext también ofrecemos un modelo para dar cuenta de las palabras que evidencian un significado emocional.

Aquí termina el primer apartado en el que hemos intentado explicar las raíces cognitivas de los fenómenos que vamos a etiquetar. Estos fenómenos se han presentado desde una perspectiva puramente introspectiva e individual. Ya sabemos que el lenguaje codifica nuestra interpretación de lo real, pero no sólo. Nosotros somos observadores de un universo infinito cuya conducta será interpretada por otros que también están observando la realidad bajo umbrales de percepción similares. Ahora bien, ¿cómo podemos estar seguros de lo que el otro está interpretando? La afirmación de que para que haya comunicación es necesaria una base de conocimientos compartida ya se ha convertido en un lugar común; sin embargo, todavía no hemos conseguido sistematizar racionalmente dicha base de conocimientos, ni predecir lo que es compartido y lo que no en una interacción comunicativa. En el siguiente apartado, hablaremos de cómo los individuos se comportan para construir y mantener una realidad convencional que los estudiosos han intentado sistematizar a través de los concep-

to de **norma, regla social y marco**. Esta realidad artificial construida a través de nuestras prácticas sociales cotidianas se ha dado en llamar **la intersubjetividad**, y en ella se asientan los cimientos del significado de las lenguas.

1.3. La construcción de la intersubjetividad: los observadores observados

Cuando los cuerpos están desnudos, las miradas están vestidas.

La presentación de la persona en la vida cotidiana. E. Goffman.

Existen muchos estudios dentro de la Ciencia Cognitiva que se han ocupado del problema de la percepción del ser humano. Uno de los más conocidos, como ya señalamos anteriormente, es el de los experimentos realizados sobre la percepción del color entre pueblos con culturas y lenguas muy diferentes. La Antropología ha criticado estos planteamientos atenuando el alcance de estas conclusiones; los antropólogos alegan que, si bien es cierto que existen unos umbrales de percepción del color similares, los colores poseen diferente significado en cada cultura, de tal manera que los seres humanos razonan en su vida cotidiana con lo que ese color significa en su contexto (como por ejemplo, el uso del color negro para expresar el sentimiento de pena en nuestra cultura). Otro ejemplo sería el del significado atribuido a los números. En la cultura china el ocho es el número de la buena suerte. Esto tiene una gran trascendencia social y económica en este país. Se llega a pagar más por los números ocho de las habitaciones de hotel; así como el número de la mala suerte, el tres, no aparece en los listines telefónicos. Las culturas se materializan en normas que regulan la conducta social de los seres humanos y le confieren a los actos una interpretación que deriva en un juicio de valor positivo o negativo. Ya hemos hablado de esto en el apartado sobre las emociones. Pensemos, por ejemplo, en la palabra *adulterio*, donde una conducta sexual se interpreta negativamente a causa de unas normas culturales de origen religioso.

Además, buena parte de nuestra actividad mental se ocupa en saber cómo nos están viendo los demás. Erving Goffman, a principio de los setenta, se propone como programa de investigación la elaboración de una teoría sociológica de la comunicación interpersonal (SERRANO, 1983). Con estas palabras introduce Goffman su ya famosa obra *La presentación de la persona en la vida cotidiana*:

En este estudio empleamos la perspectiva de la actuación o representación

teatral [...]. En las páginas que siguen se considerará de qué manera el individuo se presenta y presenta su actividad ante otros, en las situaciones de trabajo corriente, en qué forma guía y controla la impresión que los otros se forman de él, y qué tipo de cosas puede y no puede hacer mientras actúa ante ellos. (SERRANO, 1983)

Es obvio que lo importante en algunos contextos laborales, dentro de los cuales, la universidad es uno de ellos, es más importante hacer que se trabaja antes que trabajar realmente.

Las interacciones sociales constituyen, por tanto, el entramado del orden social. Goffman toma de la teoría matemática de los juegos de Von Neumann el concepto de estrategia y con él designa todos los aspectos de la interacción susceptibles de predicción, cálculo y control (SERRANO, 1983). Para dicho autor, el comportamiento humano viene regido por un conjunto de códigos, de sistemas, de reglas; y en las interacciones cotidianas se transparenta dicho sistema. El individuo debe seguir las reglas, como un ritual, si quiere continuar siendo considerado una persona normal:

Es a estas frágiles reglas y no al inquebrantable carácter del mundo externo a las que debemos nuestro indestructible sentido de la realidad. (GOFFMAN, 1959)

No es casualidad, por tanto, que Goffman utilice la metáfora del teatro a la hora de explicar cómo se comporta el ser humano de acuerdo con lo que él quiere que piensen los demás de él y con las normas que rigen el escenario en el que se encuentra. Según Wolf (WOLF, 1979), “la metáfora teatral no asegura que la vida cotidiana y el teatro sean una misma cosa, sino dice que la representación no está confinada al reino de la ficción y que constituye un dispositivo importante y esencial en la vida cotidiana [...] parece como si empleásemos nuestro tiempo no tanto en dar información como en hacer un espectáculo”. Se espera, por tanto, que cada parte reprima sus sentimientos inmediatos ofreciendo una interpretación de la situación que se considere al menos momentáneamente como aceptable por parte de los otros y del orden social. Así, si nuestro jefe es un incompetente no podremos decirle nunca lo que pensamos, bajo riesgo de que ejerza su poder echándonos del espacio laboral. Es mejor, por tanto, fingir, seguir el juego, representar teatralmente que nuestro jefe es la persona más buena e inteligente del mundo, de esta manera, no solo nuestro jefe

estará contento con la representación teatral, sino que el orden social saldrá reforzado, y la ficción, como dice Goffman, será una realidad más segura que la inestabilidad que pueda causar la propia verdad.

Pero definir una situación comunicativa significa identificar cooperativamente qué estructura da a la interacción; es muy importante aclarar que los actores luchan y cooperan para definir el sentido de la realidad del microsistema. Para ello, se ayudan de los **marcos contextuales o frames**. Un frame está conformado por las premisas organizativas de la actividad de los actores sociales; son principios de organización que gobiernan los acontecimientos. Una divergencia de frames se produciría, por ejemplo, en un robo fallido, en que el empleado del banco no tomara en serio la intimidación para que entregue el dinero. Todos los frames implican expectativas. La comunicación es una actividad posible en virtud del frame de la interacción que constituye a los sujetos en personas recíprocamente accesibles: es en las micro realidades sociales construidas por la interacción donde es posible comunicarse, es por eso que Wolf dice: “la interacción social es un circo de la conducta”(WOLF, 1979).

Así pues, siguiendo con el ejemplo del contexto laboral, dicho contexto respondería a un marco presidido por una normas que hacen que la interacción social se vuelva relativamente predecible, esto es, una interacción donde se esperan unas cosas y no otras. Por ejemplo, se espera que el jefe mande al empleado y no al revés. O que se haga caso omiso a normas más o menos explícitas como la puntualidad. En el nivel de la conversación, se esperan ciertos actos comunicativos, como por ejemplo, el saludo y la despedida, y no el insulto o la falsa acusación.

La idea de que la intersubjetividad está conformada por marcos que tipifican y hacen las situaciones predecibles no solo se abordado en la Microsociología sino también en la Psicología Cognitiva y en la Inteligencia Artificial. Sin embargo, los marcos o guiones de la IA todavía están muy lejos de implementar este conocimiento. Goffman afirma que las interacciones están sujetas a normas, a reglas o cálculos. No obstante, el conocimiento de estas parcelas no está lo suficientemente sistematizado como para implementar estas normas o reglas al mundo de la IA. En el último capítulo de esta tesis volveremos sobre esta idea de la formalización e implementación de las normas que rigen los marcos contextuales, y haremos también una primera aproximación a este tema.

Otra corriente importante dentro de la Microsociología es la **Etnometodología**. El objetivo de la misma es estudiar “los modos en que se organiza el conocimiento que tienen los sujetos de los cursos de acción normales, de sus asuntos cotidianos”(WOLF, 1979). Para los miembros de esta escuela, al igual que para Goffman, la producción de la sociedad es un esfuerzo consciente, mantenido y llevado a efecto por los seres humanos que la componen.

Es en el hacer social donde se transparentan las redes de relaciones de intersubjetividad. En este sentido, el lenguaje es el factor cimentador de la forma social. El objetivo de estudio de esta corriente sociológica es el análisis de los modos concisos en que los individuos hacen racionales y explicables sus experiencias de todos los días. Un ejemplo de estudio etnometodológico es el de D. Judnow sobre los protocolos que se siguen para gestionar la muerte en un gran hospital americano.

Pasemos ahora a la última corriente de este apartado: **el Análisis de la conversación**. Tusón nos explica cómo el término conversar viene de la palabra latina *conversari*, que significa convivir, y que a su vez se compone de *cum* y de *versari* (dar vueltas). Interesante etimología para aquellos que entiendan la conversación como un mero intercambio de información y no como un acto social, cuando no un arte. A través de la conversación entramos en contacto con el mundo, con la intersubjetividad (TUSÓN, 1997). La conversación es una habilidad socialmente reconocida y exigida. El estudiar las conversaciones significa explicar los métodos que los sujetos emplean para construir los intercambios ordenados de palabra y para manifestarse el uno al otro la naturaleza regulada, coherente, descriptiva de sus secuencias conversacionales. La conversación también está sujeta a reglas. El objetivo de esta última corriente, por tanto, consiste “en explicitar los procesamientos, reglas y modos con los cuales los locutores ordenan, construyen su propia actividad conversacional mientras la desarrollan” (TUSÓN, 1997).

La conversación se ha convertido en el centro de atención de diferentes perspectivas científicas. La Etnografía de la comunicación, la Sociolingüística de la interacción, la Etnometodología, el Análisis del discurso, la Lingüística textual y la Pragmática son algunas de ellas. Estas áreas de estudio se proponen entender los factores cognitivos, socioculturales y lingüísticos que intervienen en cualquier interacción verbal oral. La motivación teórica fundamental por las que estas corrientes estudian la conversación se basa en la creencia de que los fenómenos sociales son del mismo orden que los lingüísticos (WOLF, 1979). Las aportaciones más interesantes han sido aquellas destinadas a crear conceptos sobre la estructura de la conversación, tales como **curso, período, tema, transacción, intercambio y movimiento** (de la escuela británica de Clouthard y Sinclair); y **la secuencia, el turno, el acto, la transición de un turno a otro, el cambio de turno, la selección del hablante, las secuencias de apertura y cierre, los pares adyacentes** (por ejemplo, *gracias, de nada* (de la escuela norteamericana) (NIETO, 1995) o el concepto de **preferencia** (GALLARDO, 1991). Cuando conversamos, existen turnos que son preferidos en la conversación, los esperables o los deseados dentro de la interacción social. Así pues, es preferible que a una invitación como *¿Te apetece ir a la feria?* se responda con una aceptación antes que con

una negativa.

Las conversaciones suelen responder a patrones estructurales:

A: ¿Te apetece dar un paseo?

B: Tengo mucho trabajo

A: Vaya, otro día.

Este ejemplo, inspirado en el de Tusón (TUSÓN, 1997) responde al esquema: pregunta - respuesta - evaluación/comentario.

Así mismo, Schegoff y Sacks (WOLF, 1979) observan en sus estudios de los años setenta que los cierres conversacionales suelen constar de cuatro partes: ofrecimiento de cierre, aceptación del ofrecimiento, despedida y cierre.

Muchos de estos conceptos sobre la estructura de la conversación han sido utilizados a la hora de etiquetar corpus como señalaremos en el capítulo seis de este trabajo.

Así pues, para terminar, quede constancia aquí de que si bien la conducta humana es difícil de predecir, no lo son los comportamientos de los individuos insertos en un contexto determinado donde ocupan un rol social que, si bien puede ser impuesto, ellos construyen y confirman cada día con sus acciones y de acuerdo con el cumplimiento de unas normas cuya explicitación en la mayoría de los casos solo se pone de manifiesto cuando se incumplen, o mejor dicho, solo pasamos a ser conscientes de ellas cuando alguno de los individuos sociales las ha roto. En cualquier caso, cuando un individuo rompe una o varias normas, se le considera maleducado, conflictivo, problemático, extravagante y dependiendo del grado de no adecuación al marco hasta puede llegar a calificársele de loco. Un individuo que no se adapte a las normas que materializan un orden social será tarde o temprano rechazado por la institución o contexto. En cualquier caso, la norma nunca será cuestionada, y mucho menos lo será el orden social en el que se sustenta, sino que será el individuo el que quede absolutamente deslegitimado por los que representen y apoyen dicho orden social.

Por tanto, y para ir concluyendo, los contextos están gobernados por reglas, y las conversaciones tienen una estructura. Todo este conocimiento es de gran utilidad para los sistemas de diálogo hombre-máquina. En el capítulo último trataremos qué grado de aplicación ha tenido este conocimiento a los sistemas de diálogo hombre-máquina. Pero ya advertimos aquí que para la formalización de los contextos no solo necesitamos que los sociólogos, antropólogos expliciten las normas sino que además necesitamos construir modelos de contextos donde las normas sigan una estructura; además debemos conseguir no solo que de unas normas se deriven otras, sino que al mismo tiempo se relacionen con los enunciados lingüísticos. Las ciencias duras siempre han desprestigiado el conocimiento de las humanidades tachándolo de vago, volátil y relativo. Sin embargo, necesitan de la ob-

servación y descripción de estos campos si quieren profundizar en la formalización de los contextos.

Pasemos ahora al último apartado del capítulo, dedicado a la Pragmática Computacional.

1.4. La formalización de la intersubjetividad: Formalismo vs. Funcionalismo y Pragmática Computacional

Así pues, ya sabemos qué hay detrás de una metáfora, de una intensificación o de un marcador discursivo como *en general*. Los enunciados lingüísticos codifican significados referentes a las acciones, la certeza del hablante sobre la verdad de los enunciados, las emociones, y las normas sociales y conversacionales, entre otras cosas.

Estos fenómenos responden a un visión funcionalista del lenguaje, según la cual, la lengua es un sistema de medios de expresión apropiados para un fin (SERRANO, 1983). Como sabemos, la visión funcionalista del lenguaje, a diferencia de la formalista, privilegia el uso por encima del sistema, invierte la relación de causalidad establecida por el innatismo chomskyano entre estructuras lingüísticas y uso, estableciendo que la estructura depende del uso, y finalmente, lo que es más importante a nuestros efectos, relaciona la lengua con su uso comunicativo y le confiere un rol de suma importancia en los procesos mentales (BURGOS, 2008).

La Inteligencia Artificial y la Lingüística Computacional, a las que tan útiles les han resultado la visión formalista del lenguaje, ya que le permitía obviar todos los fenómenos correspondientes a la actuación, a la semántica y al contexto, ya está comenzando a percibir de unas décadas a esta parte que hay una estrecha relación entre la forma lingüística, el significado y el contexto, conformado, como ya hemos explicado por normas socio-culturales que interactúan con las reglas del conversación y con las estructuras lingüísticas. En el apartado de enseñanza de segundas lenguas veremos que igual de importante es memorizar una regla que nos permitirá generar infinitas oraciones como saber qué es lo que se suele decir cuando un hablante no nativo va a pedir una barra de pan. Los sistemas de diálogo, por ejemplo, deben incorporar tanto los formalismos sintácticos como los constructos ya prefabricados que se usan en determinados contextos, y que de hecho están asociados inextricablemente a ellos, ya que todo ello en consonancia y como si de un baile se tratara, está destinado a construir la intersubjetividad.

Ahora bien, ¿pueden el pensamiento, las normas sociales y las emociones codificados en la forma lingüística ser representados por un conjunto de reglas implementables

en una máquina? (MISHKOFF, 1988) ¿Se puede predecir el comportamiento lingüístico? A. Moreno señala que el modelo computacional del lenguaje presupone una modelización matemática previa de la lengua en cuestión (MORENO, 1998). Plantearse cómo se podría modelizar y sistematizar la función del lenguaje en cada contexto es uno de los retos de la Inteligencia Artificial y de la Lingüística Computacional como ya venimos señalando a lo largo de todo este capítulo.

Los investigadores en IA han dedicado poca atención a la posibilidad de incorporar componentes emotivos a los sistemas artificiales. Los actos volitivos y de deseo por el momento no se consideran temas fundamentales de investigación en IA. Se habla mucho, no obstante, de los ordenadores como sistemas intencionales, con metas y objetivos. De momento, la mayor parte de la investigación en IA está centrada en el aspecto cognoscitivo de los sistemas artificiales. Es decir, lo relacionado con el aprendizaje, la percepción, y el razonamiento (SIMONS, 1988).

Por su parte, desde la Lingüística Computacional, se han desarrollado extensiones como la Semántica de los mundos posibles o la Lógica no monótona para solucionar aspectos como la modalidad, los cuantificadores generalizados, etc. que la Lógica de Predicados Clásica no era capaz de resolver (MORENO, 1998). Además, algunas teorías y conceptos procedentes de la Pragmática Teórica ya se han intentado modelizar e implementar. La disciplina que se ha aventurado en este propósito se ha denominado Pragmática Computacional. Veamos cuáles son sus propósitos:

Computational pragmatics is concerned with the same relations between utterances and context that are the concern of sociolinguistics and applied linguistics, but from an explicitly computational point of view. This implies in the first place a concern for how to compute the relations between linguistics aspects and context aspects. There are, evidently two sides to this. On the one hand, given a linguistic expression, the question is how to effectively *decode* those aspects of it that encode context information, i.e. how to compute the relevant properties of the context. This side of this relation is in focus when the meanings of linguistic expressions are to be computed. On the other hand, when we consider language generation, where the task is to construct a linguistic expression that encodes the context information that speaker (or writer) wants to convey, the question is how to compute the relevant properties of the linguistic expression to be generated given the relevant properties of the context. (BUNT y BLACK, 2000)

Si se nos permite esta expresión, la Pragmática Computacional actualmente es una dis-

ciplina en pañales. Jurafsky (JURAFSKY, 1995) añade que la Pragmática Computacional al igual que la Pragmática en general se ocupa de la inferencia, y a continuación señala los problemas que han recibido más atención en la comunidad computacional: la resolución de la referencia (anáfora, deixis, etc.), la interpretación y generación de actos de habla, la interpretación y generación de estructura discursiva y relaciones de coherencia, y, por último, la abducción.

Cada uno de estos objetivos se han convertido en los últimos años en respectivas líneas de investigación. Sin embargo, fenómenos como la modalización discursiva no han sido considerados, y el tratamiento de los marcadores del discurso a nivel computacional es aún, como demostraremos a lo largo de este trabajo, muy deficiente, ya que no consideran los problemas computacionales que suponen la ambigüedad categorial y la polifuncionalidad de los mismos. A medida que vayamos avanzando en los capítulos, se comentarán las investigaciones que se han realizado en lo que respecta a la implementación de la cortesía, la inducción automática de la estructura retórica de los textos a través de la identificación de marcadores del discurso y la anotación del valor ilocutivo de los enunciados.

Aquí termina el primer capítulo. Lo que hemos querido mostrar brevemente antes de enfrentarnos a la explicación de nuestro modelo de anotación pragmática ha sido la relación que guardan los fenómenos que vamos a etiquetar con otras dimensiones del ser humano de naturaleza cognitivo-social que se reflejan en el lenguaje. A continuación se expondrán más detenidamente las corrientes o escuelas de pensamiento claves en la Pragmática Teórica.

Capítulo 2

Teoría pragmática

2.1. Introducción

A pesar de que fue definida en sus orígenes como el cajón desastre de la semántica, esta disciplina ha gozado de una gran proliferación en las últimas décadas, y la producción bibliográfica en todo su conjunto que existe en torno a las diferentes cuestiones y marcos teóricos de la mismas es cuando menos inabarcable. Ya que tanto el objetivo como la novedad de este trabajo no reside en la parte de teórica, voy a tocar los temas más frecuentes y las teorías más relevantes de esta disciplina pero de una manera breve y concisa.

2.2. Introducción a la Pragmática

2.2.1. Raíces filosóficas del movimiento

En un sentido amplio, pragmatismo o filosofía pragmática hace referencia a una concepción filosófica que defiende no solo una distinción entre teoría y práctica, sino también la primacía de la segunda frente a la primera (MARCONDE, 2000). En cuanto a corriente filosófica, el pragmatismo se originó a finales del siglo XIX y se desarrolló a lo largo del XX, fundamentalmente en EEUU. Sus principales representantes son C. P. Peirce, W. James y J. Dewey. En la filosofía alemana, los representantes más importantes del pragmatismo filosófico son J. Habermas y K.O. Apel, los cuales se centrarán en el análisis de las condiciones que posibilitan la comunicación, de sus presupuestos y sus implicaciones, abarcando en este programa de investigación los campos de la ética y la política. (MARCONDE, 2000)

De las vertientes que unen esta corriente filosófica, destacaremos dos puntos comunes a todas. El primero de ellos se centra en la importancia que se le concede a la práctica, a la

experiencia concreta, a los aspectos aplicados del conocimiento y a los contextos concretos de uso. El segundo punto en común de estas vertientes como corriente filosófica es el concepto de verdad. La pragmática critica la correspondencia realista entre las proposiciones y la realidad descrita por dichas proposiciones. Cuestiona la posibilidad de verdades absolutas, concebidas independientemente de un contexto. Además, complementa la noción de verdad con las nociones de suceso y felicidad, que deben dar cuenta específicamente de la consideración del lenguaje como acción, atendiendo a sus efectos y consecuencias en contextos determinados (MARCONDE, 2000).

2.2.2. Definiciones del término Pragmática

Dentro de la Lingüística, la definición más antigua de esta disciplina nos llega de la mano de C. Morris en 1938, que además fue el primero en usar este término con su uso contemporáneo. Según este autor, la Pragmática es el estudio de la relación entre los signos y sus intérpretes. Posteriormente, R. Carnap, lógico y filósofo de la ciencia, la definiría como el estudio de la relación entre el lenguaje y sus usuarios (MARCONDE, 2000).

A partir de aquí, surgen un gran número de definiciones; en el sustrato de cada una de ellas hay una parcelación diferente del significado y, por tanto, una concepción distinta de los límites de la Pragmática, de los de la Semántica y del papel que juega la forma lingüística.

Graciela Reyes, en su manual para alumnos universitarios, señala que “la pragmática empieza con un intento de encontrar el sentido de la conducta lingüística”. Y la define como “la disciplina lingüística que estudia cómo los hablantes interpretamos enunciados en contexto”.

Para un lector especializado, daré otra definición de Pragmática que me posiciona más dentro de lo que hemos llamado antes la delimitación de los límites de la disciplina y su relación con lo estrictamente lingüístico.

En el ya clásico manual de Pragmática de Levinson (LEVINSON, 1989), se presenta todo un capítulo dedicado a comentar críticamente las diferentes definiciones que existían hasta la fecha de este término. Para este trabajo, se ha seleccionado la definición que está más acorde con la concepción de Pragmática que subyace a este trabajo:

La pragmática es el estudio de aquellas relaciones entre el lenguaje y el contexto que están gramaticalizadas, o codificadas en la estructura de una lengua. (LEVINSON, 1989) pág. 8

Tal y como explica este autor en su crítica a dicha definición, aquí solo se tratan “los aspectos de la pragmática que atañen a la estructura lingüística, pero no la parte que atañe a los principios de uso del lenguaje, o al menos sólo indirectamente en la medida en que influyen en la organización lingüística” (LEVINSON, 1989). En efecto, este trabajo concentra su atención en esta relación entre contexto y forma lingüística codificada, ya que en él se trabaja con datos procedentes de un corpus cuyo tratamiento pretende ser lo más automático posible.

2.2.3. Relación entre Gramática y Pragmática, Semántica y Pragmática, y Pragmática y otras ciencias

Primero de todo, debemos decir que la bibliografía en torno a esta cuestión es muy extensa, lo cual indica la enorme preocupación que existe dentro de esta comunidad científica por delimitar estos campos; hecho que, dicho sea de paso, puede tener sus consecuencias en la financiación de proyectos de investigación y en la representación a nivel departamental de esta disciplina dentro del darwiniano mundo de la universidad. No obstante, advierto al lector de que pasaré superficialmente sobre esta cuestión, ya que en este caso, como en la mayoría, es imposible enunciar una opinión si previamente uno no se posiciona, ideológicamente hablando, dentro de una corriente determinada.

Respecto de **las relaciones entre Gramática y Pragmática**, de nuevo también en esta ocasión, los marcos teóricos nos ofrecen un amplio abanico de posibilidades. Una de las cuestiones más relevantes que se trata es la de si considerar a la Pragmática como un módulo de la gramática, a la altura de la sintaxis, el léxico o la semántica, o bien, considerarla, una dimensión de estudio, como de hecho son partidarios conocidos pragmatistas en España como Portolés, Reyes y Fuentes (FUENTES y ALCAIDE, 2002) (PORTOLÉS, 2004).

En lo que respecta a esta cuestión, nos gustaría destacar aquí que los últimos estudios desarrollados en Neurociencia defienden que, en términos biológicos, hay aspectos del lenguaje que van más allá de la gramática y que implican a otras regiones del cerebro diferentes a las tradicionalmente asociadas al lenguaje (Broca y Wernicke), como por ejemplo, las zonas dedicadas al razonamiento social y personal o a la cognición social, donde en concreto se concentra la habilidad para hacer hipótesis sobre las intenciones y las disposiciones de los otros. Así pues, las áreas del cerebro identificadas por los neurólogos para las funciones sociales pueden ser las mismas que se usan a la hora de decidir qué decir y cómo decirlo en una situación comunicativa (SCHUMANN, 1999).

En cuanto a **la relación entre Semántica y Pragmática**, insistimos en que, dependien-

do del marco teórico al que nos acojamos, la posición cambiará. En principio, la postura clásica, establecida por Gazdar en su famosa ecuación (**pragmática es igual al significado menos condiciones de verdad**) “la pragmática se ocuparía de los sentidos contextualmente condicionados que se añaden al significado puramente lingüísticos fruto de la gramática de una lengua” (PORTOLÉS, 2004). Pero a partir de aquí hay posiciones que cada vez proponen versiones más abarcadoras de la disciplina. Leech define la Pragmática como el estudio del significado de los enunciados, y a la semántica, como el estudio del significado de las oraciones. Normalmente se utiliza el término **Pragmática radical (Radical Pragmatics)** para designar a aquellos pragmatistas que defienden que muchos fenómenos que hasta la fecha se habían considerado semánticos en el fondo eran de naturaleza pragmática. La teoría de la Relevancia, que pasaremos a explicar un poco más abajo, defiende que la Pragmática se ocupa de la capacidad de la mente que permite crear significados a partir del contexto. En España, José Portolés y Martín Zorraquino ((MARTÍN y MONTOLÍO, 1998) y (PORTOLÉS, 2004)) defienden una Pragmática integrada dentro de la semántica, y se acogen a los descubrimientos de la teoría de la argumentación para justificar que “la continuación de un discurso no se debe exclusivamente a motivos contextuales o de conocimiento del mundo, sino que está lingüísticamente condicionada”. Así por ejemplo, tal y como explica Portolés en (PORTOLÉS, 1992), los operadores meliorativos y peliorativos tales como *desgraciadamente* o *afortunadamente*, determinan la consecución de los enunciados posteriores. De modo que, siguiendo un ejemplo suyo, un enunciado como *Rodríguez solo es director general* tiene un abanico amplio de inferencias, cuyas posibilidades de interpretación se ven reducidas si añadido un operador del tipo: **Gracias a dios**, *Rodríguez solo es director general*. Según este autor, el uso de este operador **fuera un contexto**, un marco de interpretación, de tal manera que, como veremos más adelante, solo unos conocimientos implícitos y no otros podrán aplicarse.

Por último, encontramos las concepciones de la Pragmática procedentes de la Enseñanza de Lenguas Extranjeras; la cuales amplían tanto el objeto de análisis de la Pragmática que, en palabras de J. Portolés, terminan por convertirla en una disciplina “extremadamente abarcadora”. Veamos cuál es la posición de este autor al respecto:

En ciertos ámbitos, se ha favorecido la identificación del estudio de la competencia comunicativa de una lengua con la Pragmática y se habla de competencia pragmática. El problema que presenta para la Lingüística esta competencia pragmática reside en que muchos fenómenos que le interesan a la competencia comunicativa tienen más que ver con el antropólogo o el sociólogo que con el lingüista. Si se amplía la pragmática hasta identificarla con el estudio de la competencia comunicativa, creo que habría que dar un paso atrás y

distinguir entre una pragmática propiamente lingüística y otra que fuera objeto de sociólogos, psicólogos o antropólogos.

Los fenómenos que vamos a etiquetar en nuestro corpus se situarían en la interfaz entre Semántica y Pragmática. Algunos de ellos, como la evidencialidad, entrarían, académicamente hablando, dentro de la Semántica; no obstante, la Pragmática también es una parcela de investigación donde se estudian fenómenos de los que la Gramática tradicional no había dado cuenta hasta el momento como, por ejemplo, la explicación del significado de ciertas partículas como las interjecciones, o los marcadores del discurso. En este sentido, los pragmatistas coinciden generalmente en señalar dos ideas, de las cuales han surgido terrenos de investigación propios y característicos de la pragmática. Dichas ideas son, en primer lugar, la crítica al generativismo y al estructuralismo —los cuales la acusaban al principio de usar criterios pretendidamente extralingüísticos —por quedarse en el nivel de la oración y explicar el código pero no el uso de la lengua, y en segundo lugar, la crítica a la Gramática tradicional en general y a la Semántica en particular, por ignorar fenómenos del significado como la modalidad, la enunciación o la polifonía (FUENTES, 1997b) .

2.2.4. Conceptos de Pragmática: enunciación y enunciado, inferencia, presuposición, implicatura y contexto

2.2.4.1. Enunciado y enunciación

La teoría de la enunciación, basada en las ideas de Emile Benveniste y enriquecida posteriormente por O. Ducrot, centra al enunciado y no a la oración en el punto de mira análisis lingüístico, ya que estudia todos los fenómenos derivados del análisis de la lengua cuando se emite en un tiempo y espacio determinado y por un hablante en concreto. El estudio del hombre en la lengua, en palabras de Benveniste (FISHER, 2004), es la perspectiva que enmarca estos trabajos de corte neoestructuralista. En este sentido, la enunciación se entiende como un proceso, mientras que el enunciado se entiende como un producto. Además, es importante destacar que los enunciados no tienen por qué ser oraciones de sujeto y predicado siempre. En C-Oral-Rom hay enunciados que son un sintagma o simplemente un adverbio, como por ejemplo, *sí*, o un signo paralingüístico como un asentimiento.

2.2.4.2. Inferencia

El estudio de las inferencias ha adquirido tanta relevancia que actualmente se consideran el núcleo de la comprensión e interpretación de la realidad y, por tanto, uno de los pilares de la cognición humana. Desvelar las claves del funcionamiento de las inferencias permitiría comprender mejor el funcionamiento mental, la adquisición y comprensión del conocimiento; al tiempo que nos aclararían las posibles conexiones que enlazan diferentes recursos cognitivos tales como razonamiento, la percepción, la memoria y el aprendizaje.

Dentro del estudio de las inferencias, el discurso se ha convertido en un objeto de estudio por parte de la disciplina la Psicología del texto, ya que este último probablemente sea el principal medio de que nos servimos las personas para expresar nuestro conocimiento.

Son muchos los intentos que se conocen por clasificar las inferencias. Se han hecho tipologías atendiendo a su contenido, a su función, a su forma lógica o a su dirección (inferencias hacia atrás e inferencias hacia adelante), siendo esta última una de las clasificaciones más utilizadas. En las inferencias hacia atrás se establece una conexión o puente entre dos frases uniendo la información recientemente leída con la inmediatamente anterior. Con mucha frecuencia se asigna a este tipo de inferencias la responsabilidad de mantener la coherencia local del texto; las inferencias hacia delante permiten predecir los hechos o los acontecimientos posteriores. Por último, está la dicotomía entre inferencia semántica e inferencia pragmática: las primeras se basan en la aplicación de reglas formales que conducen al cien por cien de la certeza, mientras que las segundas se basan en las creencias de las personas, que asumen como cierto y probable verdades que no tienen por qué serlo (LEÓN, 2003). Pongamos a continuación un ejemplo de cómo funciona esta capacidad inferencial en el ser humano.

Con motivo de la huelga de transportistas del verano del 2008, se escucharon en los medios de comunicación enunciados de este tipo: *Han matado a un piquete informador. La policía ha pasado a la acción*. Si fuéramos una especie de otro planeta con inteligencia para estudiar cómo funciona la mente y el lenguaje en los humanos, la primera sorpresa que se llevaría nuestro extraterrestre es que después de emitir este ejemplo que acabamos de señalar la mayoría de la población oyente seguramente haya inferido que es la policía quien ha matado al piquete informativo, aunque en ningún sitio del ejemplo aparezca especificada esta información, y a pesar de que nadie se atrevería a argumentar explícitamente que así ha sido a la vista de pruebas que proporciona el enunciado. Sin embargo, esto habrá sido lo que probablemente se haya inferido como una verdad, y a pesar de que dicha verdad diste mucho de la que es producto de un silogismo o de un razonamiento lógico. Esto es lo que se ha denominado inferencia. Las inferencias es el centro de análisis de teorías que posteriormente trataremos como la de Grice o la de Sperber y Wilson.

Siguiendo con nuestro imaginario extraterrestre, la segunda cosa que seguramente le sorprendería bastante es la de que las inferencias a pesar de ser implícitas sean compartidas por oyente y hablante. Veamos qué nos dice el investigador José Portolés al respecto en su artículo *Pertinencia y Pragmática* (PORTOLÉS, 1994):

El considerar las inferencias como un constituyente de la comunicación obliga a pensar que algún mecanismo permite que sean previsibles; de otro modo los hablantes solo constituiríamos enunciados que no exigieran ningún enriquecimiento inferencial, por temer a no ser comprendidos.

Aunque la telepatía sea cosa de los mundos de ficción, buena parte de la comunicación se realiza dando por supuesto que el otro sabe lo que nosotros estamos pensando. De tal manera que existe una relación proporcional en las relaciones humanas entre comunicación y conocimientos compartidos, y por tanto inferencias comunes.

2.2.4.3. Presuposición e implicatura

La pasada Semana Santa interrumpí mi animada conversación con un gran amigo para pedirle fuego a uno de los dos barrenderos que pasaban por allí. Como el primero me contestó que no, le pregunté al compañero, y este, que tampoco tenía, el lugar de contestarme *yo tampoco*, sus conocimientos de la lengua española le hicieron responder: *yo también*.

Creo que este es un buen ejemplo para comenzar a hablar del concepto de presuposición a aquellos que no conozcan el significado de esta palabra. Parece claro por tanto que por presuposición se entiende aquella información en la que la proposición del enunciado se apoya. En el caso arriba señalado, el enunciado *yo, también* se apoya en la presuposición de no tener fuego, que por ser negativa requeriría la palabra *tampoco*, en lugar de la de *también*, regla que el hablante no nativo todavía no había adquirido.

La presuposición es uno de los temas predilectos de la Pragmática pero, sin embargo, no es la única disciplina que la toma como objeto de estudio. Respecto de la presuposición existen dos grandes debates. El primero de ellos gira en torno a su naturaleza gramatical; el segundo, en torno a su naturaleza semántica o pragmática. Respecto del primero, Chomsky ya se posicionó en su día diciendo que la presuposición formaba parte de la sintaxis ya que es necesaria para aclarar asuntos que él considera de naturaleza sintáctica como, por ejemplo, la elipsis (ALVAREZ y GINORIA, 1996). La semántica lógica, por su parte, considera que es a ella a quien les corresponde estudiar este concepto y su realización en la lengua, ya que esta información ni es de naturaleza contextual ni depende del hablante.

Así por ejemplo, en el famoso y citadísimo ejemplo de Frege: *El rey de Francia es calvo*, se presupone que Francia tiene rey, sin recurrir al contexto. Sin embargo, y ya para terminar, la Pragmática considera que la presuposición también es su responsabilidad ya que hay inferencias que sí se basan en el contexto y en la posición del hablante respecto de lo enunciado, que serían aquellas que se basaran en aquellos conocimientos de naturaleza cultural o social y no en la información gramatical que se acaba de recibir, la cual se va archivando a medida que se va recibiendo y que tiene consecuencias en la gramática de los enunciados que vienen después, como el caso de la elipsis en el ejemplo de *yo, también*.

Para terminar con el tema de la presuposición, me gustaría hacer solamente un último comentario en torno al tratamiento que se le da en los estudios que he consultado al respecto. En ellos, las presuposiciones son representadas gráficamente en términos de proposiciones lógicas, y sus estudiosos dan por supuesto que de hecho las presuposiciones son proposiciones lógicas, cuando todavía desconocemos empíricamente qué forma de representación mental tiene el conocimiento inferido, tanto el presupuesto como el implicado, del cual pasaremos a hablar a continuación.

Si la presuposición podría definirse como una inferencia que funciona hacia atrás, la implicatura se puede definir como una inferencia que funciona hacia adelante. Un ejemplo de implicatura lo tenemos en el caso que mencionamos más arriba sobre las diferentes conclusiones que se pueden obtener de enunciados del tipo *Es pobre pero honrado*: me casaré con él; *Es honrado pero pobre*, no me casaré con él.

Para terminar este apartado me gustaría volver a la cuestión de la predictibilidad de las inferencias. Tanto de aquellas que funcionan hacia atrás (presuposición) como las que funcionan hacia adelante (implicatura). ¿Cómo es posible que infiramos cosas que no están dichas claramente y que la mayoría de las veces, aunque no todas, acertemos? A pesar de que desarrollaremos más esta reflexión en el capítulo dedicado a la Inteligencia Artificial, me gustaría ya adelantar que tal y como ha señalado Morgan (en (ESCANDELL, 1996)) “los mecanismos inferenciales pueden sufrir procesos de institucionalización”, es lo que Morgan llama, recurriendo a una metáfora biológica, **implicaturas cortocircuitadas**, esto es, dinámicas de pensamiento o, como a mí me gusta llamarlas, rutas neuronales tan trilladas que se vuelven evidentes, tan evidentes que no hace falta explicitarlas. El *perdona pero no quise decir eso*, por ejemplo, es una fórmula que en muchos casos se usa para atenuar estas inferencias convencionalizadas cuya institucionalización escapa a nuestro control y uso individual de la lengua.

Para ilustrar esto último, me gustaría incluir el magnífico ejemplo que G. Reyes incluye en su manual de Pragmática para explicar el concepto de Pragmática:

Porque te quiero mucho dijo Ludmilla, y por una de esas astucias del idioma el mucho le quitaba casi toda la fuerza al quiero.

Julio Cortázar, *Libro de Manuel*, 1973, pag. 94. en (REYES, 1990)

2.2.5. El concepto de contexto

¿Qué es el contexto? Por un lado, hay un acuerdo general en subrayar la importancia de este concepto para el estudio del uso del lenguaje:

La concepción inferencial de la comunicación solo se puede explicar si se tiene en cuenta el contexto. Este contexto siempre es mental y lo forma un conjunto de suposiciones que permiten la comprensión de un enunciado. Estas suposiciones o se hallan ya en nuestra memoria, o se crean en nuestra mente en el momento de la comunicación [...] Para dar cuenta del uso efectivo de una lengua, es tan importante la propuesta de un principio que guíe las inferencias como la explicitación del ordenamiento de nuestro contexto mental. (PORTOLÉS, 2004)

A pesar de dichas afirmaciones sean ampliamente compartidas, no se ha llegado todavía a un acuerdo sobre cómo racionalizar, sistematizar y en definitiva dar cuenta de los contextos. Lo más cercano a dicha sistematización, muy pobre a efectos pragmáticos, han sido los guiones (secuencias de actos estereotipadas basada en nuestra memoria de situaciones) programados desde la IA, desde otras parcelas como la Pragmática o la F. del Lenguaje, se intentan establecer primero qué tipo de conocimientos conformaría el contexto, pero todavía nadie ha logrado formalizar en su sistema de reglas las normas y los conocimientos que están detrás de una situación comunicativa. Llevar a la práctica esta idea es todo un reto intelectual sobre todo si tenemos en cuenta la visión de la **la Teoría de la Relevancia**, según la cual, el contexto es algo dinámico:

Cualquier suposición nueva se procesa siempre en relación con el contexto mental que proporcionan otras suposiciones anteriores, las cuales funcionan a modo de premisas en la operación inferencial de interpretación. Según este planteamiento, en un proceso inferencial, la información conocida opera como contexto a partir del cual se contrasta la información nueva. (MONTOLÍO, 1998)

En tercer lugar, el contexto desempeña un papel decisivo en la interpretación de todos los enunciados. En esta teoría el contexto no está predeterminado, se construye al interpretar. Para Sperber y Wilson, el contexto es un conjunto de premisas usadas en la interpretación de un enunciado; es decir, una serie de suposiciones que son accesibles mentalmente para los interlocutores. Así pues, el contexto sería de naturaleza cognitiva e implicaría creencias, recuerdos, conocimientos del mundo, etc. Cualquier suposición nueva se procesa siempre en relación con el contexto mental que proporcionan otras suposiciones anteriores, las cuales funcionan a modo de premisas en la operación inferencial de interpretación. Según este planteamiento, en un proceso inferencial, la información conocida opera como contexto a partir del cual se contrasta la información nueva. (MONTOLÍO, 1998).

A esta concepción del contexto como algo flexible, algo que se va rehaciendo conforme se produce la interacción, ha contribuido la obra del sociolingüista John Gumperz. Gumperz propuso el concepto de **índices de contextualización** para referirse a distintos elementos lingüísticos o no lingüísticos que sirven, si bien muchas veces de forma inconsciente, para construir el contexto en una situación determinada, y que además influyen en cómo se comprende el mensaje. Estos índices contribuyen a la contextualización necesaria para que los hablantes interpreten adecuadamente la actividad discursiva que se está realizando y actúan, al menos, en los siguientes niveles: la prosodia, los elementos paralingüísticos, la elección de código y la elección de formas léxicas o de fórmulas. La deixis social también sería una índice de contextualización (PORTOLÉS, 2004).

De nuevo como en el caso de la presuposición, existen muchas dificultades no solo para definir qué incluye el contexto, sino también para saber cuál es la mejor forma de representarlo. La **Psicología cognitiva** propone que los contextos son modelos mentales (GIVON, 2001), y desde las vertientes más lingüísticas (KARTTUNEN, 1977) se defiende que estos se materializan en formas lógicas.

El debate sobre el conocimiento implícito, ya se hable de contexto, implicaturas, preposiciones o de inferencias en general, siempre nos devuelve al problema de la interacción entre el lenguaje natural y el lenguaje del pensamiento. La Lingüística Computacional y la IA tienen que contar con este saber implícito tan importante para la comprensión del funcionamiento de las lenguas naturales como instrumentos de comunicación.

Pasemos pues, una vez explicados los conceptos básicos de esta disciplina, a la exposición de sus marcos teóricos más relevantes, ya que la Pragmática a día de hoy no dispone de un modelo teórico unificado. Lo que el lector encontrará aquí será una breve exposición de cada una de ellas. Cada uno de los planteamientos que presentamos a continuación aportan

una visión nueva de la comunicación lingüística en el ser humano, y bajo mi punto de vista ¹, estos conocimientos se tienen que tener en cuenta a la hora de elaborar una teoría general de la comunicación.

La exposición parte de las teorías de naturaleza cognitiva, continúa con las que tienen un origen social, y termina con las que se centran más en la estructura lingüística. Los enfoques se presentarán por este orden: Teoría de la Relevancia, Teoría de los Actos de Habla, Máximas Conversacionales y Principio de Cooperación de Grice, Cortesía Lingüística, y, por último, Teoría de la Argumentación, un orden bastante diferente al cronológico².

2.3. Pragmática y cognición: La teoría de la Relevancia

Tal y como ya se apuntó en la introducción, comenzamos con la Teoría de la Relevancia porque, a diferencia de las demás, esta teoría ofrece una gran amplitud explicativa sobre el funcionamiento de la comunicación en el ser humano.

Creo que para un lector no especializado, la mejor manera de entender lo que propone en términos básicos la teoría de la Relevancia es imaginando lo que nuestra mente hace cuando nos encontramos delante de las manchas que se utilizan en los test de Rorschach; esas manchas que en principio no significan nada y que son moldeadas por cada persona, de acuerdo con lo que hay en su mente, hasta hacerlas relevantes, hasta conferirles un significado.

La Teoría de la Relevancia de Sperber y Wilson presenta un modelo de comunicación alternativo al tradicional basado en la codificación-descodificación de signos lingüísticos. En el nuevo modelo, ³ durante la comunicación, tienen lugar los siguientes procesos: un proceso de interpretación de la realidad, una codificación lingüística de lo más relevante, una descodificación del mensaje y una interpretación del mismo en la mente del interlocutor.

Los procesos de codificación y de enriquecimiento pragmático están regulados por un principio cognitivo, esto es, automático e inconsciente, denominado Principio de Relevancia, según el cual, el hablante solo codifica la parte más relevante de lo que quiere co-

¹Aviso al lector en este punto de que usaré la voz de la primera persona del plural para las cuestiones relativas a la narración, mientras que usaré la voz en primera persona del singular para las cuestiones relativas a la opinión, como bien se puede ver en este caso.

²El orden cronológico es: Teoría de los actos de habla (Austin y luego Searle), Principio de Cooperación e inferencias de Grice, Horn, Brown y Levinson, y Sperber y Wilson.

³Sperber y Wilson en realidad son, hablando figuradamente, hijos de Grice, y este ya entendía el proceso de comunicación no como un proceso de codificación y descodificación, sino como un proceso de inferencia. (PORTOLÉS, 1993)

municar, mientras que el oyente infiere todo lo que el hablante ha querido decir partiendo solamente de la parte codificada verbalmente por el hablante. La comunicación, por tanto, es inferencial, y está regulada por la ley de coste y beneficio, según la cual siempre se minimiza el coste de procesamiento y se maximiza el beneficio en términos de conocimiento. Solo se transmite lo que es relevante, y algo que es relevante es una información que despierta inferencias, que despierta significado a todos los niveles en las bases de conocimiento de los participantes.

Esta teoría se encuadra dentro del paradigma de las Ciencias Cognitivas, que siempre ha manifestado su interés por el modo en que el ser humano procesa y representa la información en el interior de la mente. En dicho paradigma, la mente se entiende como un mecanismo eficiente de procesamiento de la información (MONTOLÍO, 1998), como un dispositivo mecánico destinado a consumir información (somos informavívoros en palabras de Miller) (BONILLA, 1996) y que se hace comprensible desde un punto de vista científico usando el ordenador como metáfora estrella. Además, dicha teoría no sólo es respetuosa con el modularismo de Fodor sino que, frente a su pesimismo, se propone abrir un camino a la descripción del procesador central. La pragmática quedaría fuera por tanto de la competencia lingüística del hablante, no formaría parte del módulo lingüístico dedicado en la arquitectura diseñada por Fodor, y donde el generativismo ha encontrado un sitio para integrar su teoría dentro de dicha arquitectura.

La comunicación debe ser definida, como consecuencia, como un proceso inferencial. Esto no sería posible sin la capacidad del ser humano para hacer inferencias o razonamientos. No es necesario explicitar lo que es fácilmente inferible por el otro. En la fraseología del español encontramos bastante unidades fraseológicas con contenido metafórico que aluden al procesamiento inferencial de la información: *ser cortito, no llegar, no coger los chistes, no pillar nada, etc.* En el plano sociocultural, existen maneras diferentes de hacer inferencias según los marcos de cada uno. Así por ejemplo, las famosas diferencias sobre el uso del lenguaje por parte de las mujeres y los hombres remiten a procesos de interpretación diferente de las formas lingüísticas en un contexto determinado, y por tanto, a diferencias de criterio a la hora de establecer qué información es relevante y cuál no.

Por lo tanto, debemos empezar a comprender los enunciados lingüísticos como índices que conducen a un significado que no está en los textos y que está relacionado con la intención del hablante, pero que al mismo tiempo se apoyan en la base de conocimientos comunes para asegurarse de la interpretación correcta del enunciado.

Así pues, según los partidarios de esta teoría, el proceso de descodificación o recepción pasa por diferentes etapas (si son seriales o paralelas se desconoce por completo): desambiguación en la asignación de referentes, enriquecimiento o especificación de referencias

(explicaturas ⁴), y conclusiones (implicaturas).

Un mensaje relevante es un mensaje que crea efectos contextuales, que enriquece el mundo de conocimientos del que lo recibe sobre la base de sus supuestos. No siempre los mensajes tienen el mismo grado de relevancia, puede ser que la información nueva sea tan nueva que sea incapaz de relacionarse con algún supuesto y de integrarse en el mundo de creencias; en otras ocasiones, puede ocurrir que la información para el locutor sea conocida y por tanto menos relevante; y finalmente también puede ocurrir que sea tan contradictoria que sea difícil atribuirle un significado. En principio, lo ideal sería que el coste fuera mínimo y el beneficio máximo, pero pensemos que el aprendizaje, por ejemplo, de una lengua extranjera requiere en su primera etapa un coste procedimental alto frente al beneficio recibido en términos de información relevante.

En el proceso de interpretación de un enunciado, buena parte de su relevancia pasa por inferir su relación entre el valor de verdad y el plano de la acción social, como lo que sucede cuando decimos *Me estás pisando*.

La teoría de la Relevancia ha recibido bastantes críticas en lo que respecta fundamentalmente a su estatus científico. Como ya hemos señalado, esta teoría defiende que la comunicación se rige por un principio cognitivo que automáticamente enriquece los enunciados escuchados con el conocimiento almacenado en la mente creando implicaturas. Durante esta operación se buscan para los enunciados los mayores efectos contextuales, mayor beneficio en términos de conocimiento con el menor coste de procesamiento. Cuatro son las críticas que vamos a enumerar a continuación.

La crítica más importante a esta teoría, dentro del marco popperiano, es que no se puede falsar, esto es, no se puede demostrar que sea falsa, ya que todavía no podemos medir el coste de energía de procesamiento cognitivo que realiza un hablante al interpretar un enunciado (PORTOLÉS, 1994); respecto de esta crítica, Montolío (MONTOLÍO, 1998) señala que esto “se trata de un problema empírico que ha de resolver la neurociencia, no la pragmática”; sin embargo, desde mi punto de vista, aunque no sea competencia de la Pragmática calcular empíricamente el coste de procesamiento de la información, no debemos de obviar que esta teoría se levanta sobre unos valores de verdad que hoy en día, a pesar de las técnicas de la Neurociencia para visualizar la actividad cerebral, son muy difíciles de demostrar, y tendrán que pasar aún unos años antes de saber si al principio de Relevancia se le puede considerar una entidad relevante dentro del funcionamiento de la mente entendida como una entidad física. De momento, la mente es como la energía oscura sabemos que

⁴las explicaturas de bajo nivel incluyen presuposición, anáfora, deixis, las de alto nivel incluyen la actitud del hablante respecto de lo enunciado (ESCANDELL, 1996)

está allí, que influye en los procesos de creación de la materia pero no podemos dar cuenta de ella.

La segunda crítica, enunciada por el psicólogo Sánchez Zavala, está dirigida a la consideración de la comunicación como un proceso exclusivo de intercambio de información, negando la parte emocional de dicho proceso (MONTOLÍO, 1998).

En tercer lugar, la teoría de la Relevancia explica cómo los seres humanos procesamos la información dentro del marco de la comunicación, sin embargo, este marco no explicaría el funcionamiento del lenguaje interior dentro de la mente, y su función cognitiva (MONTOLÍO, 1998).

En cuarto lugar, algunos autores alegan que esta visión de la Pragmática entraría dentro de la Psicología más que de la Lingüística o de la Gramática, ya que deja fuera de su programa de investigación el estudio sobre los usos sociales de la lengua (PORTOLÉS, 2004).

2.4. Teoría de actos de habla

2.4.1. Conceptos Básicos

La Teoría de los Actos de Habla, enunciada en sus orígenes por el filósofo del lenguaje Jean Austin y completada posteriormente por John Searle, es una de las reflexiones existentes más sistemáticas sobre la relación entre el uso cotidiano del lenguaje y el plano de la acción humana. Según Austin, el lenguaje ha sido a lo largo de la historia un instrumento que el hombre ha ido perfeccionando de generación en generación con el fin de servir eficazmente para todos los usos de la vida. Los enunciados no solamente expresan acciones sino que sirven para realizar acciones. Así por ejemplo, el acto de casarse se materializa cuando el hablante dice *Sí, quiero*. Otro ejemplo, muy citado en la bibliografía (PORTOLÉS, 2004) es el de la realización del acto de divorcio en algunos países del mundo árabe, el cual se lleva a cabo diciendo tres veces el enunciado *Me divorcio de ti*. Por tanto, y contrariamente al modo en que se venían estudiando los enunciados, no sólo es importante saber si un enunciado es verdadero o falso, sino que también es necesario saber si el acto que profiere es adecuado o no a una situación concreta, ya que esto asegurará la felicidad del mismo. Para que un enunciado se convierta en un acto tienen que darse una serie de condiciones tales como la existencia de un procedimiento convencional, la existencia de personas y circunstancias apropiadas, una actuación determinada por parte de los participantes, la sinceridad de los mismos, etc. Siguiendo con el ejemplo de la boda, sería un infortunio el hecho de que el que casara fuera en lugar de un cura el cantante de AC/DC; esto en pragmática se llamaría, como decimos, infortunio, e invalidaría el acto que

hay detrás del *Sí, quiero*. Por otra parte, si existe una falta de coincidencia entre lo que el enunciado dice que hace y lo que en realidad hace se llama infelicidad. Por ejemplo, que alguien te trate de insultar con la palabra *maricón* y la persona no se sienta insultada, al no considerar esta palabra un insulto sino como un adjetivo puramente descriptivo de una propiedad de la persona.

Para Austin, hablar una lengua es tomar parte en una forma de conducta gobernada por reglas. El lenguaje es un tipo particular de acción, y debería integrarse dentro de una Teoría general de la Acción; según este autor, un estudio de las propiedades exclusivamente formales de una lengua no tendría sentido si no va acompañado de un estudio sobre el papel que desempeñan tales formas (ESCANDELL, 1996). Al igual que Goffman, Austin coincide en señalar que toda actividad lingüística es convencional, esto es, está controlada por reglas y los principios reguladores que utiliza el lenguaje no son, en esencia, diferentes de los que se sigue en todas las actividades humanas.

En el significado de un enunciado, hay que valorar tres tipos de actos cuyo conjunto se ha denominado **la tricotomía de Austin**:

1. **Acto locutivo**: el acto realizado por el mero hecho de decir algo. Por ejemplo, *Hay que bajar la basura*, es una oración afirmativa cuyo significado es la suma de sus partes.
2. **Acto ilocutivo**: el acto que se realiza al decir algo. En este caso, el acto ilocutivo del enunciado *Hay que bajar la basura* podría ser en un determinado contexto, una orden.
3. **Acto perlocutivo**: el acto que se realiza por haber dicho algo, los efectos producidos. Siguiendo con el ejemplo, el acto perlocutivo sería que el receptor del enunciado terminara bajando la basura.

Así pues, un enunciado tendría un significado, una fuerza y unos efectos, y el acto de habla sería la unidad mínima de la comunicación lingüística.

Los estudiosos del campo de la Teoría de Actos de Habla suelen afirmar que Searle lleva a sus últimas consecuencias las ideas de Austin (ESCANDELL, 1996). Dos son las aportaciones más importantes de Searle a los planteamientos de Austin. El primero es asumir la idea de que realizar un acto ilocucionario es comprometerse a una forma de conducta gobernada por reglas. Una vez hecha esta asunción, Searle se propone reflexionar

sistemáticamente sobre dichas reglas y su relación con el contenido proposicional de los enunciados. Para ello, este autor diferencia entre **reglas regulativas** y **reglas constitutivas**. Un ejemplo de la primera sería *el melón se corta con cuchillo*. Las reglas constitutivas serían por ejemplo las reglas del ajedrez o de cualquier juego, donde **X cuenta como Y**. Así pues, de la misma manera que un *touch down* cuenta como seis puntos, una promesa cuenta como una obligación futura. Por tanto, los actos ilocucionarios son actos realizados de acuerdo a un conjunto de reglas constitutivas, que de acuerdo con Searle la comunidad científica debe dejar claramente definidas como lo están las reglas del ajedrez.

La segunda novedad respecto de la teoría de su predecesor es la afirmación de que existe una relación regular y constante entre la fuerza ilocutiva y la forma lingüística. Searle presenta los siguientes tipos de actos de habla:

1. **Asertivos**, que indican cómo son las cosas.
2. **Directivos**, que nos dicen qué cosas hacen.
3. **Compromisivos**, que encierran la promesa de hacer cosas.
4. **Expresivos**, que expresan sentimientos y actitudes.
5. **Declarativos**, que producen cambios a través de nuestros enunciados.

Otra idea interesante para la Lingüística Computacional es el intento de formalización de la teoría. Teniendo en cuenta el tipo de acto de habla, la formalización de un enunciado con su fuerza ilocutiva sería la siguiente: $F(PR)$. Donde: F representaría los valores ilocutivos, P el predicado y R el sujeto.

No obstante, este es uno de los puntos más controvertidos de su posicionamiento teórico. Cada uno de los tipos de acto de habla según Searle está asociado con una determinada estructura lingüística. Los indicadores de la fuerza ilocutiva serán:

1. Curva de entonación
2. Énfasis prosódico
3. Orden de palabras
4. Predicados realizativos

Esta es la tipología más seguida en las aplicaciones de esta teoría a la Inteligencia Artificial. Sin embargo, cuando la teoría de los actos de habla se usa en la enseñanza de segundas lenguas los actos de habla se multiplican sin orden ni concierto. ¿Cuántos actos de habla hay en realidad? ¿Es una categoría cerrada o abierta? Austin insiste en su obra *Cómo hacer*

cosas con las palabras en que la tipología de los actos se puede describir de la misma manera que el zoólogo clasifica una especie determinada. Todavía no hay un acuerdo en torno a estas preguntas, ni tampoco sobre la diferencia entre las acciones propias de los actos ilocutivos como *pedir perdón*, *aseverar* o *preguntar* y el resto de las acciones humanas como *amar*, *leer* o *cocinar*.

Para ir ya terminando me gustaría añadir que se ha contraargumentado a este autor con multitud de ejemplos en los que un mismo enunciado dependiendo del contexto puede tener un acto ilocutivo u otro. Escandell se pregunta en su manual si los investigadores en la materia han interpretado bien a Searle. En cualquier caso, las críticas que se le han hecho no están fundamentadas empíricamente ya que no hemos llegado a crear sistemas de descripción y de transcripción tan finos de los patrones de entonación como para comprobar si tiene o no razón, y de lo mismo se puede decir de la descripción del énfasis y del orden de palabras. Fenómenos registrados por la Lingüística actual pero que están muy lejos de ser descritos con detalle suficiente como para llegar a enunciar reglas de formalización en relación a los significados que codifican.

Como señala G. Reyes (REYES, 1990) “la teoría de los actos de habla presupone en efecto que todo acto lingüístico revela una intención y que es un ejercicio de sinceridad; pero postula a la vez que todo acto de habla para ser realizado también debe valerse de fórmulas reconocibles, que circulan en nuestro tesoro lingüístico”. Por lo tanto, necesitamos fórmulas reconocibles, repetibles para hacer promesas, pedir perdón, elogiar, insultar, etcétera. Además, al igual que cuando hablamos de las inferencias, debemos añadir que los significados ilocutivos, esto es, las acciones implícitas en los enunciados, también están sometidas a convención. Como consecuencia de ello, habrá enunciados cuyo significado ilocutivo esté absolutamente convencionalizado, como por ejemplo en el caso de *¿Me puedes pasar la sal?*, y otros en los que la interpretación sea más libre y más dependiente del contexto: *Tengo frío*. Morgan señala que hay dos tipos de convención, las convenciones lingüísticas y las sociales. El enunciado *¿Me puedes pasar la sal?* sin llegar a ser una expresión idiomática está codificando en su proposición una convención de uso social.

Me gustaría concluir la exposición de los puntos más relevantes de esta teoría destacando la importancia de los significados ilocutivos en los enunciados, estos son tan imprescindibles en la comunicación que tal y como señala José Portolés (PORTOLÉS, 2004) a veces se nos olvida lo que nos han preguntado pero, sin embargo, no se nos olvida que era una pregunta.

2.5. El contrato comunicativo: Grice, Levinson, Lakoff

El centro de la investigación de Grice respecto de las oraciones está en el significado del hablante. “Al proferir una emisión, un hablante intenta comunicar algo y, a la vez, intenta que su intención comunicativa se reconozca por un oyente: intenta, por ejemplo, inducir en él una creencia o lograr que se lleve a cabo determinada acción mediante el reconocimiento de su intención (la del hablante)”(GRICE, 1991); lo que se busca en definitiva es que el oyente intente hacer algo o de que éste crea que el hablante cree algo.

Grice centra buena parte de su obra en explicitar los principios que regulan los intercambios comunicativos y que influyen en la interpretación de los enunciados. De nuevo, el significado implícito o inferido de un enunciado se convierte en el objeto de atención también por parte de este filósofo. Según este autor, los intercambios comunicativos están regulados por el Principio de Cooperación ⁵, el cual, a su vez, se desglosa en cuatro categorías: **cantidad, cualidad, relación, y modalidad**. La paráfrasis de dichas categorías en normas sería: **dé usted la cantidad de información necesaria, no mienta, sea relevante, y sea claro**, respectivamente. A diferencia del Principio de Relevancia, **el Principio de Cooperación** es puramente descriptivo, y se considera una manifestación más entre otras de la conducta racional del ser humano. A pesar de que Grice no aventura en ningún momento de su obra que dicho principio sea de carácter cognitivo, sí reconoce que no solo el comportamiento lingüístico se puede explicar mediante dicho principio, sino también cualquier tipo de actividad de la conducta humana que esté basada en la cooperación. El principio de racionalidad es una especie de “condición preparatoria que se espera que los hablantes observen”(ESCANDELL, 1996).

Una de las aportaciones más importantes de este filósofo es su famosa clasificación de las inferencias que provocan los enunciados y que no vamos a detallar aquí. Según él, en un enunciado hay que diferenciar por un lado lo que se dice, lo cual sería evaluable por una lógica de tipo veritativo-condicional, y por otro lado, lo que se comunica, esto es, el contenido implícito de un enunciado y que él denomina implicatura. Para ser más exactos, la implicatura sería la recuperación de un nuevo significado después de aplicar las reglas de deducción al enunciado escuchado. Dentro de las implicaturas, se hace hincapié en las conversacionales que surgen como resultado de la violación de algunas de las máximas y del supuesto de que se respeta el principio de Cooperación. ⁶

⁵Haga usted su contribución a la conversación tal y como lo exige, en el estadio en que tenga lugar, el propósito o la dirección del intercambio que usted sostenga (GRICE, 1975)

⁶Recomendamos el manual de V. Escandell (ESCANDELL, 1996) a aquellas personas que quieran profundizar en la taxonomía de implicaturas elaborada por este filósofo del lenguaje.

Grice reconoce que “el esquema necesita ser ampliado para producir propósitos generales como influir y organizar las acciones de los otros” (ESCANDELL, 1996). Leech amplió este esquema con la máxima *sea cortés*, mientras que Sperber, Horn y Levinson lo redujeron, como se explicará en la siguiente sección.

Los postulados conversacionales de Grice han recibido **críticas** por parte de la Etnografía. Ochs (OCHS, 1977) cuestiona todas las máximas conversacionales de Grice a través del estudio de una comunidad lingüística que habita en Madagascar. En dicha cultura, la gestión de la información nueva y conocida al igual que las normas de cooperación son diferentes. Así por ejemplo, se evita por todos los medios el uso de deícticos de persona ya que albergan la creencia de que esta práctica atrae fuerzas enemigas. Además si alguien tiene una información nueva que el resto de la comunidad desconoce pero que sabe que existe, la persona en cuestión automáticamente adquiere prestigio dentro de la comunidad y se espera que dicha persona no revele la información. Por último, evitan sistemáticamente realizar afirmaciones explícitas acerca de creencias, actividades y comprometerse con hechos futuros para no asumir la responsabilidad social que conlleva.

2.6. La vertiente social de la comunicación

2.6.1. Los seguidores de P. Grice

Pasemos ahora a revisar los modelos que intentan dar forma a la dimensión social de la comunicación, vertiente que se denomina Pragmática Social y, más concretamente, Cortesía Lingüística. Cuatro autores son tradicionalmente señalados en la bibliografía: R. Lakoff, G. Leech, P. Brown y S. Levinson. Estos autores han pasado a la historia de la Pragmática por el intento de formalizar en términos de reglas, principios y fórmulas, respectivamente, el escurridizo mundo de cómo los seres humanos gestionan en sociedad las relaciones humanas con el uso de la lengua. Al mismo tiempo, una característica común que une estos tres modelos es el deseo de completar el modelo conversacional de Grice. No está de más recordar que el principio de cooperación tiene como meta asegurar una transmisión de información eficaz; partiendo de este supuesto, la cortesía es una estrategia al servicio de las relaciones sociales; por tanto, en un extremo está la transacción y en el otro la interacción. Este es el modo en que lo han enfocado estos estudiosos: las relaciones humanas siempre son el resultado de una tensión constante entre el intercambio de información y el cuidado de los intereses sociales. Por tanto, para estos autores, la Cortesía, entendida como el conjunto de estrategias o normas sociales orientadas a regular la relación social (agresividad, respeto, cooperación, competitividad) de los miembros de una comu-

nidad, sería el factor determinante por el que el Principio de Cooperación no se aplicara en toda su efectividad.

Por último, tampoco está demás señalar que la cortesía forma parte intrínseca de la racionalidad humana, la cual es por encima de todo una racionalidad social, siempre tenemos que atender a las normas, incluso para saltárnoslas, y siempre debemos considerar que la sociedad en realidad da muy poco espacio para decir a nivel cotidiano lo que uno realmente piensa.

El primer intento por sistematizar las normas de cortesía es elaborado por R. Lakoff desglosa la máxima de sea cortés en tres modalidades: **no se imponga, ofrezca opiniones y refuerce los lazos de camaradería.**

G. Leech, por su parte, eleva lo que para Lakoff era un regla al estatus de principio: **el Principio de Cortesía**, el cual se puede dividir a su vez en seis máximas: **tacto, generosidad, aprobación, modestia, acuerdo y simpatía.** Dichas máximas regulan respectivamente el coste y el beneficio del hablante y del interlocutor en términos de autoridad, beneficio, desprecio, aprecio, desacuerdo y simpatía. Este principio, “permite evaluar el grado de adecuación entre el enunciado y la distancia social” de los participantes (ESCANDELL, 1996). Aquellos actos que exigen más coste para el destinatario (y por tanto susceptibles de ser descortesés) son más susceptibles de indirección, y como consecuencia, se sacrifica el principio de eficacia informativa. Veamos un ejemplo de indirección del mensaje para atenuar el daño en el segundo enunciado:

1. Tu hijo se ha suicidado cortándose las venas.
2. Su hijo podría haber fallecido por una herida de arma blanca.

Por lo tanto, para Leech el principio de cortesía sería un principio regulador de la conducta.

2.6.2. Un modelo universal de cortesía

El modelo de Brown y Levinson constituye un paso más en la sistematización del modo en que se gestionan las relaciones sociales. Escandell (ESCANDELL, 1996) afirma que es el intento más elaborado y estructurado de explicar los motivos de la cortesía en las lenguas, al tiempo que es un buen punto de partida para los interesados en estudiar la relación entre

estructuras o forma lingüística y estructuras sociales. Para nosotros, uno de los atractivos de este trabajo es su pretensión de universalidad. (ESCANDELL, 1996)

Al igual que Leech, también parten de la metáfora del coste-beneficio para explicar las relaciones sociales, pero enriquecen su modelo con algunos conceptos de la obra de Goffman, como por ejemplo, el de **territorialidad** ⁷ y el de imagen.

El concepto de imagen es universal, todo ser humano que viva en sociedad tiene una imagen pública y buena parte de sus conductas o de sus comportamientos están orientados a conservarla y cultivarla. Esta imagen pública tiene dos facetas: una positiva y otra negativa. Con la **imagen positiva**, los autores aluden a la necesidad que todo ser humano que viva en sociedad tiene de sentirse aceptado por el entorno social en el que habita. Por **imagen negativa**, en cambio, se entiende la necesidad que todo ser humano tiene de sentirse libre. La acción, o mejor dicho, las acciones que se espera que se realicen mediante la enunciación lingüística (por ejemplo, en el enunciado *Me estás pisando*) pueden afectar, dañar esta imagen. Para evitarlo se despliegan toda una serie de estrategias sociales y lingüísticas que intentan atenuar el coste social para el destinatario ⁸. Estos autores proponen una fórmula para calcular el riesgo social que para la imagen pública puede tener un acto considerado amenazante:

⁷El territorio propio incluye todo lo que tiene que ver con el ámbito de actuación del individuo, a saber, sus sentimientos y pensamientos, cuestiones personales e íntimas y los territorios temporal, corporal y espacial (CARRASCO, 1999).

⁸Las diferentes estrategias pueden clasificarse en dos tipos: cortesía negativa y positiva. La cortesía negativa sería aquella cortesía que intenta no invadir el territorio de su interlocutor. La cortesía positiva invade el territorio del interlocutor en un afán de mostrar interés, preocupación y cariño por la persona. Hay personas, clases sociales, países y culturas en las que hay un predominio de un tipo de cortesía por encima de la otra. Lo que para una persona con cortesía positiva es una muestra de cariño o solidaridad para una persona con cortesía negativa es directamente una intromisión y un abuso de confianza. Y viceversa, alguien con cortesía positiva puede interpretar que los de cortesía negativa son personas secas y egoístas.

$$\text{Riesgo (AAIP)}_x = (D+P+G)_x$$

Tabla 2.1: Fórmula para calcular el riesgo social de la Teoría de la Cortesía

AAIP	Acción que amenaza la imagen pública
D	Distancia social
P	Poder
G	Grado de imposición del acto

Tabla 2.2: Explicación de las signaturas de la fórmula de arriba

En función de estos factores, los hablantes construyen enunciados que evidencian estrategias sociales que oscilan entre estos dos parámetros: estrategias abiertas/encubiertas y estrategias directas/indirectas. Un ejemplo de una combinación de los dos parámetros, una estrategia abierta e indirecta, sería el enunciado *¿Me prestas mil pesetas?*

Según el modelo de Brown y Levinson, la justificación última de estas conductas lingüísticas es el hecho de que toda sociedad necesita controlar y canalizar la agresividad de sus miembros (ESCANDELL, 1996). A mi modo de ver las cosas, no solo se trata de agresividad, sino también de reprimir, gestionar o controlar otras emociones como la expresión o exteriorización de la euforia, la pasión, el cariño, la inseguridad o el miedo.

Este modelo ha recibido muchas **críticas** generalmente orientadas a cuestionar su universalidad. Ha habido muchas aportaciones, procedentes fundamentalmente de la Antropología (DURANTI, 1997), que critican el modelo presentado por estos autores por estar concebido dentro y para una cultura anglosajona (etnocentrismo). De nuevo, nos enfrentamos a la eterna diatriba entre universalismo y relativismo cultural. Desde mi punto de vista, esta teoría es universal en sus fundamentos teóricos: todo el mundo tiene una imagen pública y todas las relaciones sociales hasta las más íntimas o altruistas o puras, pueden explicarse atendiendo a los parámetros de coste y beneficio social. Dicho esto, es obvio que cada sistema u organización socio-económico-cultural regula a su modo el coste y el beneficio de los grupos sociales o individuos que componen una sociedad, y sus conductas solo son explicables dentro de los parámetros que rigen su entorno.

Es terreno de la Sociología, la Antropología y la Etnometodología establecer y explicar las reglas, principios o normas de cortesía en las que se apoya un orden social determinado; pero, en mi opinión, es oficio de la Pragmática y de la Sociolingüística establecer este puente entre lo social y la forma lingüística, en el caso de la cortesía que está codificada verbalmente. En este sentido, como ya se aclaró en la introducción, defendemos una postura absolutamente funcionalista: si una comunidad considera relevante para sus estruc-

turas sociales algo, tiende a marcarlo o a convencionalizarlo lingüísticamente, como por ejemplo, el sistema de honoríficos del japonés.

La Pragmática Computacional vería con muy buenos ojos cualquier tipo de conocimiento formalizado procedente de este campo, incluso si este conocimiento es relativo y se restringe a una cultura, país o contexto situacional determinado.

2.7. Teoría de la Argumentación de Anscombe y Ducrot

Dejando al margen de este trabajo una gama muy variada de teorías de la argumentación (para un breve resumen de las mismas recomendamos (GILLE, 2001)) nos centraremos en este apartado en la Teoría de la Argumentación de Anscombe y Ducrot.

Para estos autores, la argumentación es un acto cotidiano. Todo acto de lengua argumenta, es más, el uso de la lengua ya es un acto de argumentación en sí mismo que trata de convencer al interlocutor de lo que nuestra subjetividad está percibiendo. Por lo tanto, constantemente estamos argumentando frente a otros y frente a nosotros mismos nuestra visión del mundo. Para este fin, utilizamos un gran abanico de estrategias lingüísticas que operan a nuestro favor y a través de las cuales guiamos u orientamos las inferencias de nuestros interlocutores hacia las conclusiones deseadas. El objetivo de esta teoría consiste en analizar el modo en que los enunciados se ligan entre sí en favor de determinadas conclusiones.

Una de las características más importantes que debemos reseñar y que ha tenido consecuencias muy importantes en las motivaciones de este trabajo ha sido la idea defendida por estos autores de que el proceso inferencial está marcado o codificado verbalmente por la propia estructura lingüística de los enunciados, y no sólo por su contenido. Como se señala en (ESCANDELL, 1996), *la Teoría de la Argumentación se ocupará de los medios formales que proporciona la lengua a sus hablantes para orientar argumentativamente sus enunciados; y, paralelamente, de los medios formales que sirven al destinatario para construir su interpretación.*

¿Cómo podemos aprovechar estos conceptos para una Pragmática Computacional? Según estos autores, la Lógica Clásica no puede dar cuenta de los siguientes fenómenos. Veamos el siguiente ejemplo:

1. ¿Está ya la cena?

2. Sí, casi.

La secuencia *Sí, casi* incurriría en una contradicción lógica, y sin embargo, desde un punto de vista argumentativo, es perfectamente aceptable.

En la argumentación lingüística, a diferencia de la argumentación lógica, *ni el número de argumentos está predeterminado de antemano, ni la conclusión es necesaria y automática* (ESCANDELL, 1996). V. Escandell ejemplifica esta idea con los siguientes ejemplos:

1. No tengo ganas de seguir estudiando: voy a dar una vuelta.
2. No tengo ganas de seguir estudiando, son las ocho de la tarde, y además, hace calor: voy a dar una vuelta.
3. No tengo ganas de seguir estudiando, son las ocho de la tarde, y además, hace calor: me prepararé una copa bien fría.

Observemos ahora estos dos ejemplos también tomados de la misma fuente:

1. # María es muy lista: habla inglés, chino e incluso francés.
2. María es muy lista: habla inglés, francés e incluso chino.

Según Escandell (ESCANDELL, 1996), para Anscombe y Ducrot es la diferencia en la estructura lingüística, y no en el contenido, lo que explica el contraste entre los dos últimos ejemplos. Para comprender lo que ocurre en estos enunciados Anscombe y Ducrot han acuñado una serie de conceptos como los de **orientación, suficiencia, peso y escala argumentativa, y topoï** que configuran el núcleo duro de su investigación.

Los enunciados tienen una orientación y un peso argumentativo, esta es la razón por la que el enunciado *María es muy lista: habla inglés, chino e incluso francés* suena raro. Los argumentos pueden organizarse en escalas argumentativas como esta, donde el último argumento introducido por la palabra *incluso* debe tener un peso argumentativo mayor que los anteriores. En (2) saber francés no tiene más peso que saber chino. Saber inglés, francés y chino constituye un conjunto de argumentos, una clase argumentativa en la terminología de Ducrot, que están coorientados a una misma conclusión, la de demostrar que María es

lista. Para que esto se cumpla, la clase argumentativa debe apoyarse en un razonamiento argumentativo, un tópico o lugar común que estos autores han llamado **topos o topoi**. En este caso, el topos sería el siguiente: *cuantos más idiomas se sabe, más listo se es*. Un topos es, pues, una regla general, que se supone aceptada por el sentido común, que tiene un carácter gradual, y que es cancelable, es decir, uno siempre puede argumentar que no es necesario saber idiomas para ser listo.

La última aportación importante que nos gustaría señalar aquí es la que estos autores han hecho a la hora de profundizar en la semántica de los elementos extraoracionales. Entre ellas, está su diferenciación entre conector y operador discursivo. El criterio de clasificación sería el alcance que el marcador tiene sobre los enunciados entre los que se encuentra. Según J. Portolés (PORTOLÉS, 1998a), un **conector argumentativo** es una unidad que articula dos miembros del discurso (o más) que intervienen en una estrategia argumentativa única. Ejemplos de conectores discursivos son: *pero, sin embargo, además, etc.* Por otro lado, un **operador argumentativo** es una unidad que, aplicada a un contenido, transforma las potencialidades argumentativas de este contenido. Ejemplos de operadores son: *bien, casi, un poco, etc.* Muchos operadores forman parte del sistema gramatical, e institucionalmente no se consideran marcadores del discurso; por ejemplo, el adverbio *mero* en el enunciado *Ha sido un mero accidente*.

Además, estos autores clasifican los conectores en función de tres criterios: función, valencia y fuerza argumentativa.

1. **Función:** si el enunciado es argumento o conclusión
2. **Valencia:** el número de argumentos que enlaza
3. **Fuerza orientativa:** orientados y antiorientados.

Respecto de las relaciones entre Semántica y Pragmática, otro de los temas predilectos de esta teoría, Anscombe y Ducrot proponen una Pragmática integrada en la Semántica, que solo podrá ocuparse de todo aquello que derive de las propiedades del sistema mismo. La Pragmática así concebida no podrá tratar todo lo que es situacional o conversacional, porque estos son siempre fenómenos que involucran necesariamente elementos extralingüísticos. Sin embargo, como apunta V. Escandell (ESCANDELL, 1996) *no acaba de estar del todo claro que los topoi no sean ya elementos ajenos al sistema de la lengua: los tópicos que sirven de base a la argumentación son una buena muestra de ello*.

Para terminar me gustaría hacer una reflexión sobre cómo integrar estos conceptos dentro de una Pragmática Computacional. El concepto de topos es una idea que no debe de ser

desdeñada por la IA a la hora de crear modelos de razonamiento entre conocimiento verbalizado y conocimiento implícito. Ahora bien, ¿qué disciplina debe encargarse de explicitar los topoi? ¿La Pragmática, la Semántica, la Microsociología, la Inteligencia Artificial en la elaboración de sus sistemas de expertos o el Análisis crítico del discurso que se encarga de analizar la ideología que hay detrás de los discursos? No cabe duda de que su explicitación es necesaria para la comprensión de los enunciados y la formalización del sentido común, pero la cuestión es cómo hacerlo. Volvamos al ejemplo de los idiomas que sabe María. Desde un punto de vista computacional, hasta donde hemos desarrollado las herramientas de análisis sintáctico, estas estructuras son exactamente iguales, ¿cómo reconocer y comprender la diferencia de aceptabilidad o de adecuación entre un enunciado y otro? Para ello deberíamos introducir bases de conocimiento estructuradas en escalas argumentativas, donde se reflejara que saber chino ocupa en la escala argumentativa un lugar más alto que saber francés dentro de la cultura occidental; y que los argumentos que se acompañen por la palabra *incluso* deben tener un peso argumentativo mayor que los anteriores.

Un buen comienzo sería empezar por localizar y clasificar argumentativamente los marcadores del discurso. Estos nos darían pistas sobre la orientación argumentativa y el peso, como por ejemplo, el caso de *incluso*. Y en este sentido, este trabajo pretende realizar una investigación básica en lo que a la recolección, descripción y anotación en corpus de estas partículas se refiere.

2.8. Conclusiones finales

Aquí termina el capítulo dedicado a la exposición de las teorías más importantes del campo de la Pragmática. En España, tenemos representantes de todas las corrientes teóricas presentadas aquí. Destacan especialmente en Teoría de la Argumentación, José Portolés y Marta Tordesillas, en la Universidad Autónoma de Madrid, y Catalina Fuentes en la Universidad de Sevilla. Los que más han trabajado bajo el marco Teoría de la Relevancia han sido Estrella Montolío y Salvador Pons Bordería (del grupo Valesco), en Barcelona y Valencia respectivamente. Henk Haverkate, Victoria Escandell y Diana Bravo han desarrollado un papel fundamental y digno de ser mencionado en el cultivo de los estudios de cortesía en lengua española, un terreno apenas explorado hace tan solo unas décadas. En la divulgación al alumnado universitario de los conocimientos de esta disciplina en lengua española debemos destacar la labor de investigadores como Graciela Reyes, Victoria Escandell, José Portolés o Marta Tordesillas. Gran parte de los conocimientos aquí expuestos de una manera mucho más sintética proceden de sus manuales como habrá podido notar el

lector especializado en la materia. Antes de pasar al siguiente capítulo me gustaría poner de manifiesto algunas críticas que ha recibido la Pragmática como disciplina.

Ya he tratado durante la introducción y el primer capítulo algunas cuestiones relativas al estatus científico de la Pragmática. Entre los comentarios que he realizado, me gustaría volver a insistir en la falta de preocupación por la formalización. A G. Reyes le gusta definir la Pragmática con el adjetivo de empírica, y maneja la oposición entre empirismo y formalismo, oponiendo el amor de la Pragmática por los datos, por el desprecio de los generativistas por los mismos. Citando sus palabras: “La Pragmática cada vez es más empírica, se acerca a la Sociolingüística y deja de preocuparse por la influencia filosófica, por los principios abstractos y por su formalización”. A propósito de la formalización de la pragmática, R. Cerdá señala las críticas de Chomsky a cualquier línea de investigación destinada a la definición del significado de los enunciados a partir de la intencionalidad del emisor en cada acto comunicativo, no porque Chomsky considere que dichos intentos no estén fundamentados sino porque los considera, en palabras de Cerdá (CERDÁ, 1995), “científicamente inaccesibles, a causa de la imprescriptibilidad que rodea a la mayoría de los ingredientes esenciales a todo acto comunicativo.” Aparte de su imprescriptibilidad, la ausencia de unidades discretas con las que trabajar, algo que a la Lingüística tan buenos resultados le ha reportado, se considera como una crítica más al cientificismo de esta disciplina.

En tercer lugar, a pesar de que, como ya hemos señalado, parece predominar la postura de que gramática y uso se complementan, como señala G. Reyes (REYES, 1990) “todavía no se ha llegado a una teoría aceptable sobre la relación entre la forma y la función, entre lo que la lengua codifica en su gramática y lo que los hablantes cumplen.”

En cuarto lugar, una de las críticas también más sonadas acerca de esta materia es la falta de integración teórica (REYES, 1990). Si resumimos los tipos de Pragmática, podríamos distinguir:

1. Una **Pragmática Formal**, que defiende la resolución de sus problemas con los modelos matemáticos y lógicos ya aplicados a la semántica formal (Editorial, 2007) y que también incluiría teorías dinámicas como Discourse Representation Theory (DRT), de Kamp y Reyle (1993).
2. Una **Pragmática Cognitiva o Cartesiana**, basada como diría V. Escandell (ESCANDELL, 2004) en factores internos.
3. Una **Pragmática Holística**, que abraza todos los aspectos contextuales de la comunicación lingüística: mental, social, cultural, físico, interaccional e histórico.

4. Una **Pragmática Lingüística**, que en palabras de J. Portolés (PORTOLÉS, 2004), explique que la palabra *encantado* solo se dice la primera vez que te presentan a alguien, o que se ocupe de que una misma persona sea llamada según el contexto: *presidente Berlusconi o papi*.

Para concluir, diremos que la asignatura pendiente de esta disciplina es superar el debate entre Pragmática Cognitiva y Pragmática Social, e integrar estas dos dimensiones en un modelo general que vaya desde la cognición hasta la forma lingüística. Victoria Escandell, una de las investigadoras más importantes en este terreno, ha hecho alguna propuesta para unificar en un modelo general la Pragmática Social y la Pragmática Cognitiva (ESCANDELL, 2004). Según esta autora, la primera debe ser concebida en normas, y la segunda en principios (ESCANDELL, 2004). Además de estas posturas, debemos destacar puntos comunes de los planteamientos teóricos expuestos como el uso de la metáfora del coste-beneficio en la teoría de la Relevancia y en la de la Cortesía, o la consideración del significado procedimental de los marcadores del discurso tanto por parte de la teoría de la Relevancia como de la Argumentación. Sin embargo, debería haber acuerdos sólidos en torno a la definición, delimitación del campo y herramientas conceptuales básicas. Solo así, podrá ser útil a otras parcelas de conocimiento como la Psicolingüística, por ejemplo, en el estudio del autismo, la Enseñanza de segundas lenguas, o la Lingüística Computacional.

Pasemos ahora al capítulo tercero de este trabajo, dedicado a la categoría de análisis por excelencia de los estudios pragmáticos: los marcadores del discurso.

Capítulo 3

Los marcadores discursivos en las teorías pragmáticas

3.1. Introducción

Con este capítulo entramos en el área de estudio de los marcadores del discurso, uno de los objetos de investigación más importantes no solo para la Pragmática sino también para otras disciplinas afines a ella como la Gramática del Texto, el Análisis Crítico del Discurso o el Análisis Conversacional.

El capítulo está dividido en dos grandes apartados. En el primero de ellos se tratan cuestiones teóricas básicas en torno al concepto y delimitación gramatical de la categoría pragmática **marcador discursivo**. En la segunda parte, se expondrán, primero, las aportaciones más relevantes que las diferentes escuelas lingüísticas han hecho a la hora de dar cuenta del significado y funcionamiento de estas formas lingüísticas, y segundo, las aplicaciones de estos conocimientos al mundo de la Lingüística Computacional y de la Inteligencia Artificial.

3.2. Definición de marcador discursivo y características gramaticales

La definición oficial dentro del ámbito académico hispano sobre marcador discursivo es la siguiente:

Los marcadores del discurso son unidades lingüísticas invariables, no ejercen una función sintáctica en el marco de la predicación oracional - son, pues, elementos marginales -, y poseen un cometido coincidente en el discurso: el de

guiar, de acuerdo con sus distintas propiedades morfosintácticas, semánticas y pragmáticas, las inferencias que se realizan en la comunicación (MARTÍN y PORTOLÉS, 1999).

A continuación, enumeraremos las características lingüísticas más importantes con las que se ha delimitado el concepto de marcador discursivo dentro de la gramática y las cuales han sido recogidas fielmente del capítulo dedicado a marcadores del discurso en la *Gramática Descriptiva* de I. Bosque y V. Demonte (MARTÍN y PORTOLÉS, 1999):

1. Son unidades invariables.
2. Poseen diferente grado de gramaticalización.
3. Tienen una alta movilidad sintáctica, siempre que se trate de categorías mayores.
4. Están delimitados por marcas entonativas.
5. Carecen de la posibilidad de recibir especificadores y adyacentes complementarios, aunque hay excepciones como, por ejemplo, *además* en *además de X*.
6. No se coordinan entre sí, pero se pueden coordinar con sintagmas que se sitúan en posición de inciso en el caso de los adverbios que funcionan como marcadores.
7. No admiten negación.
8. Cuando se encuentran dentro una unidad sintáctica, la relación sintáctica que guardan es con la totalidad del sintagma.
9. No pueden ser destacados por perífrasis de relativo.
10. Tienen diferentes grados de independencia con la oración en la que aparecen; algunos incluso pueden aparecer solos.
11. Presentan diferente grado de autonomía.
12. Pueden situarse en miembros del discurso que constituyen categorías léxicas y sintagmáticas muy diversas, no solo oraciones. Ejemplos de J. Portolés y M.M. Zorraquino (MARTÍN y PORTOLÉS, 1999):

a) *Ilusión, imaginación y, en definitiva, arte.*

b) *Además, europeos.*

c) *O sea, mal.*

d) *En todo caso, sinceridad.*

Estas son las características lingüísticas más importantes de los marcadores discursivos. Como ya se ha indicado durante su enumeración, no todos los marcadores del discurso cumplen todos los requisitos que se acaban de mencionar. Como bien ha puesto de manifiesto la Gramática Cognitiva (CUENCA y HILFERTI, 1999), los marcadores del discurso también están sometidos a los procesos de gramaticalización y lexicalización propios de todas las lenguas en evolución. Esto hace que nos encontremos ante una clase bastante heterogénea y difícilmente estabilizable por criterios puramente formales. Así por ejemplo, el criterio de la invariabilidad es bastante relativo. J. Portolés (MARTÍN y PORTOLÉS, 1999) señala explícitamente que debido a la aplicación de este criterio, se dejan fuera de este concepto algunas partículas como *por este motivo, por eso, por ello, por esta razón, etc.* Este es uno de los grandes problemas a la hora de decidir qué es marcador discursivo y qué no. M. M. Zorraquino y J. Portolés (MARTÍN y PORTOLÉS, 1999) han establecido la invariabilidad como un parámetro discriminador. Nosotros no vamos a seguir este criterio. En este sentido, convenimos con las últimas investigaciones por parte de la Lingüística Cognitiva en materia de marcadores del discurso en que estas partículas, al igual que otras categorías, pueden analizarse desde los parámetros de periferia y prototipicidad. Así por ejemplo, un reformulador prototípico sería *o sea* o *es decir*; mientras que *o si se quiere, si se prefiere, o lo que es lo mismo, si se prefieren otras palabras, dicho de otra forma, dicho de otra manera, dicho de otro modo, esto quiere decir que, quiero decir*, serían ejemplos de marcadores discursivos menos gramaticalizados, y por tanto, se situarían, al igual que los pingüinos cuando hablamos de la categoría de aves o pájaros, en las periferias de la categorización de la realidad. Con lo cual, en principio proyectaremos nuestro modelo de anotación pragmática¹ sobre aquellas formas lingüísticas que actúan en las zonas externas de la oración, y que por tanto no cumplen un papel temático dentro de la estructura eventiva del verbo. Esta es más o menos el grueso de palabras o formas lingüísticas que pretendemos abarcar, aunque como diremos más adelante, estamos intentando saber cómo determinados fenómenos se materializan lingüísticamente, con lo que aunque en principio este es nuestro radio de acción, también hemos incluido en nuestra lista palabras que forman parte de la oración como por ejemplo la formación *se dice que*, como marcador evidencial. En realidad, marcamos unidades con información pragmática, solo que las hemos decidido denominar marcador discursivo para no ampliar aún más la extensa nómina de denominaciones que este tipo de palabras ha recibido en la tradición lingüística (CORTÉS,

¹Relaciones discursivas, lenguaje emocional, modalización, evidencialidad, actos de habla y deixis

1995a) y (CORTÉS, 1995b).

Pasemos ahora a profundizar más en la función que cumplen estas partículas en el discurso a la luz de los marcos teóricos más relevantes de la Pragmática.

3.3. ¿Qué ha dicho la Pragmática de los marcadores discursivos?

3.3.1. Los marcadores discursivos: relevancia y argumentación

Desde el marco de la Pragmática, los marcadores discursivos se han visto como modificadores del significado de los enunciados. Leamos las palabras de M. Zorraquino al respecto:

El concepto de marcador discursivo remite a aquellas palabras que hacen referencia al proceso de enunciación, donde influye la posición del hablante con respecto a lo que está diciendo, la presencia de alguien que supuestamente está escuchando y la co-referencia a sendos contextos, discursivo y mental (de conocimientos, creencias, etc.), compartidos por hablante y oyente. Los marcadores del discurso se pueden ver como las huellas léxicas del propio procesamiento del discurso durante la interacción (MARTÍN y PORTOLÉS, 1999).

Los estudios de Diane Blakemore (BLAKEMORE, 1987), enmarcados dentro de la **Teoría de la Relevancia**, han puesto de manifiesto el significado procedimental o procedural (según las traducciones) de estas partículas en el discurso. Tal y como nos explica E. Montolío (MONTOLÍO, 1998), “desde esta perspectiva teórica, los conectores se entienden como pistas que el hablante utiliza a fin de dirigir cooperativamente el proceso interpretativo de su interlocutor”. Su función principal sería la de “imponer restricciones respecto a cuál debe ser el contexto en el que el enunciado que los contiene debe interpretarse, y en consecuencia, guían la trayectoria para acceder a dicho contexto, a través del reforzamiento de unas inferencias o la eliminación de otras que equivocadamente pudieran suponerse. En otras palabras, los conectores y otras partículas se conciben como instrucciones metapragmáticas para procesar la información nueva en el contexto cognitivo adecuado”.

La Teoría de la Argumentación, por su parte, también coincide con esta visión de los marcadores discursivos. Al igual que Blakemore, Anscombe y Ducrot estudian el modo en que estas partículas influyen en la consecución de los enunciados posteriores. Estos autores,

junto con la nómina de investigadores españoles que han utilizado su marco de análisis a la hora de establecer la función y el significado de estas partículas (M. Tordesillas, C. Fuentes, J. Portolés, etc.), han hecho también aportaciones teóricas interesantes, de las cuales vamos a destacar las siguientes:

1. La distinción entre los marcadores que son operadores y los marcadores que son conectores argumentativos.
2. El estudio del significado de los marcadores en función de la Teoría de los Topoi.
3. El estudio del significado de los marcadores a la luz de la Teoría Polifónica.

La primera y segunda cuestión ya se trataron en el capítulo anterior. Hagamos aquí una breve introducción a la última Teoría, la última aportación teórica de los autores de la Teoría de la Argumentación, y a la luz de la cual se han estudiado muchos marcadores del discurso en español (por ejemplo, (FUENTES, 1997a) y (PORTOLÉS, 1998b)). Nos apoyamos en J. Portolés (PORTOLÉS, 1998b) , para la explicación de la misma.

Ducrot mantiene que el hablante no ha de ser concebido como una unidad. En cada acto de habla los seres humanos ponemos en juego tres tipos de personajes: el sujeto empírico, el locutor y el enunciador. El sujeto empírico es la persona real, el locutor es el que mantiene una posición discursiva, mientras que el enunciador es el que defiende un punto de vista determinado.

Citando un ejemplo de J. Portolés, un padre y una madre pueden ser delante de su hijo dos sujetos empíricos y un mismo locutor, ya que representan delante del hijo una misma voz.

3.3.2. Clasificación de los marcadores del discurso

A continuación, presentamos una de las clasificaciones más sistemáticas que se han presentado dentro de los estudios de marcadores discursivos en español en el mundo académico hispanohablante. Existen otras tipologías como las ofrecidas por L. Cortés (CORTÉS y CAMACHO, 2005), A. Briz (BRIZ y HIDALGO, 1998) o C. Fuentes (FUENTES, 1998)², pero buena parte de los criterios de estas clasificaciones son de naturaleza estructural, esto es, clasifican el marcador discursivo en función de unidades discursivas como acto, micro-acto, secuencia, etc. La implementación computacional de esta clasificación pasaría

²Adición, oposición, causa activa, temporal, reformulación; y dentro de la reformulación: explicación, precisión, corrección, recapitulación, conclusión y ejemplificación

por la existencia de un corpus de textos anotados ya previamente en estas unidades estructurales. Sin embargo, C-Oral-Rom, como ya explicaremos en el capítulo correspondiente, solo está etiquetado en turnos y unidades prosódicas (unidades tonales y enunciados). Nos interesa por tanto la clasificación de J. Portolés (MARTÍN y PORTOLÉS, 1999) porque basa su clasificación en el tipo de operación discursiva.

Veamos la tabla 3.1 que hemos extraído de la obra de estos autores (MARTÍN y PORTOLÉS, 1999):

ESTRUCTURADORES DE LA INFORMACIÓN	Comentadores	pues, pues bien, así las cosas...
	Ordenadores	en primer,segundo lugar ...
	Digresores	a propósito ...
CONECTORES	Conectores aditivos	Además, encima, aparte, ...
	Conectores consecutivos	por tanto, en consecuencia, ...
	Conectores contraargumentativos	en cambio, por el contrario, por contra, antes bien, sin embargo, no obstante, con todo, etc.
REFORMULADORES	Explicativos	o sea, es decir, esto es ...
	De rectificación	mejor dicho, mejor aún ...
	De distanciamiento	en cualquier caso, en todo caso
	Recapitulativos	En suma, en conclusión ...
OPERADORES ARGUMENTATIVOS	Operadores de refuerzo argumentativo	en realidad, en el fondo, de hecho
	Operadores de concreción	por ejemplo, en particular, ...
MARCADORES CONVERSACIONALES	De modalidad Epistémica	claro, desde luego, por lo visto
	De modalidad Deontica	bueno, bien, vale, etc.
	Enfocadores de Alteridad	hombre, mira, oye, etc.
	Metadiscursivos Conversacionales	bueno, eh, este, etc.

Tabla 3.1: Clasificación de marcadores del discurso de J. Portolés.

Como se verá en el siguiente capítulo, en el que explicamos el modo en que hemos clasificado semánticamente los marcadores del discurso de C-Oral-Rom, convenimos con esta clasificación en su fundamentación teórica, pero de cara a la formalización tiene una serie de desventajas que dificultan su implementación computacional. Veamos por qué.

En primer lugar, en esta tipología se diferencian cuatro grandes grupos de marcadores: estructuradores de la información, conectores, reformuladores, operadores argumentativos y marcadores conversacionales. No queda explicitado si los estructuradores de la información, los reformuladores y los marcadores conversacionales son o no conectores u operadores. Por tanto, en esta clasificación, la primera gran división está expresando simultáneamente dos tipos de información: la operación discursiva (ordenación discursiva y reformulación) y la categoría pragmática (conector u operador).

En segundo lugar, se diferencian los marcadores de concreción, pero ¿dónde están los de generalización como *en general* o *generalmente* o *por regla general*? Si se distingue una operación como la concreción, por qué no se hace otro tipo para la generalización.

En tercer y último lugar, el último grupo, el de los marcadores conversacionales recoge una tipología de fenómenos: modalidad epistémica, deóntica, enfocadores de alteridad y metadiscursivos, que poco tienen que ver entre sí, y que se han puesto junto bajo un criterio, esto es, ser propios de registro conversacional/oral, que no se aplica a la hora de clasificar el resto de los tipos de marcadores. Esto es, deben aplicarse los mismos criterios o parámetros para toda la clasificación de estas partículas en familias lingüísticas. Todas las clases deberían haberse definido en función de los parámetros seleccionados: conector u operador, tipo de operación discursiva y, por último, tipo de registro; y no unos tipos en función de si es conector u operador mientras que otros, en función de una operación o registro discursivo determinado, como el último grupo. Esta cuestión de los criterios de clasificación a la hora de establecer taxonomías tiene una enorme importancia en la Lingüística Computacional, entiéndase por tanto la crítica desde esta perspectiva.

Por último, no se diferencia entre modalidad epistémica y evidencialidad. *Por lo visto* es un marcador evidencial, muy diferente a *claro* y *desde luego*, que son intensificadores de la certeza que el hablante tiene sobre el valor de verdad del enunciado.

En este trabajo se propone ir de la semántica a las categorías léxicas y además, separar cada tipo de información en un sistema rasgos. En nuestro etiquetado, habrá un rasgo para cada fenómeno, y cada rasgo tendrá sus propios valores, de esta manera, por ejemplo, habrá un rasgo que sea modalización discursiva, y sus valores serán intensificación o atenuación.

3.3.3. La Estructura textual y los marcadores del discurso

El objeto de estudio de este marco es, como su propio nombre indica, la conversación, y distingue para su análisis una serie de unidades en las cuales se estructuraría la misma. Aunque existen, por un lado, los planteamientos procedentes del marco inglés (que se ocupa de la macroestructura de los textos) y, por otro, las aportaciones de la escuela francesa (ocupado de su microestructura) (NIETO, 1995), las dos propuestas tienen en común las siguientes unidades de estructuración: **interacción, secuencia, intercambio, intervención (movimiento) y acto de lenguaje** (BRIZ y HIDALGO, 1998). En España, tenemos propuestas que han seguido estos marcos como las ya mencionadas más arriba de L. Cortés y M. Camacho (CORTÉS y CAMACHO, 2005) y A. Briz junto con su grupo de investigación VALESCO.

Para los estudiosos de la conversación cotidiana, los marcadores discursivos son partículas relevantes, ya que marcan la estructura de los textos. Según el planteamiento de Briz (BRIZ y HIDALGO, 1998), los marcadores discursivos serían tipologizados en dos grupos: los argumentativos (que son explicados desde el marco teórico de Ducrot y Anscombre) y los metadiscursivos, encargados de la actividad formulativa, esto es, de evidenciar la ordenación y organización jerárquica de la conversación. A su vez también son vistos como aportes en la cohesión y coherencia discursiva.

En cuanto al significado de los marcadores discursivos, según estas corrientes, estos variarían en función de la zona en que se encuentre dentro la conversación. Así pues, para el marcador discursivo *bueno* Briz distingue hasta siete funciones o valores (BRIZ y HIDALGO, 1998).

Otra de las corrientes lingüísticas que se ha ocupado de los marcadores del discurso ha sido **La Gramática del Texto**³. En España, uno de sus cultivadores más importantes en materia de marcadores discursivos es Manuel Casado Velarde. A continuación presentamos sus reflexiones en torno a la relación entre marcadores del discurso y funciones textuales:

El punto de contexto -o la transición, si se quiere- entre la Lingüística del
Texto y los marcadores textuales de un idioma reside en que los idiomas (al

³**Definición de Gramática del texto:** La Gramática del Texto es la disciplina que se ocupa de estudiar la relación entre funciones textuales y categorías idiomáticas que existe en las lenguas a la hora de construir textos. Por ejemplo, la pregunta es una función textual que frecuentemente se expresa mediante la categoría idiomática de entonación interrogativa, pero no solo, sino que también se expresa por medio de otras estrategias como en el siguiente ejemplo: *por favor, indíqueme dónde está la catedral*. Igualmente, una orden no necesariamente siempre se expresa mediante imperativo, como ocurre, por ejemplo, en el caso de *¿Te quieres callar?* (CASADO, 1993).

menos todos los que se conocen) disponen de estos recursos especiales, más allá de la oración, para expresar lo que venimos denominando sentidos o funciones textuales, organización del discurso, etc. Un enfoque coherente con tal comprobación debe llevar a plantearnos si, para expresar una determinada función textual, que conocemos en virtud de nuestra competencia textual, existe(n), alguna(s) forma(s), regulada(s) idiomáticamente en una lengua concreta. (CASADO, 1998)

Desde la Lingüística Textual, los marcadores discursivos se conciben como materializaciones de funciones textuales. Así por ejemplo, el español dispone para expresar la función discursiva de la **digresión**, formas como *por cierto*, *a propósito* y *a todo esto*.

La Gramática del Texto, que se apoya en buena parte de su fundamentación teórica en el funcionalismo de Halliday, ha recibido muchas críticas respecto al papel que cumplen los marcadores discursivos en su contribución a la coherencia⁴ y cohesión textuales⁵.

A pesar de todo ello, de cara a nuestra investigación, este planteamiento, que va de las funciones textuales a las formas lingüísticas, resulta atractivo de implementar a nivel computacional. No obstante, todavía en el nivel teórico, existen problemas por resolver como la creación de una tipología de funciones textuales estándar. Ya el propio Casado Velarde reconoce este problema:

Quede constancia de que no existe un catálogo exhaustivo de funciones textuales: las que aquí se citan son algunas de las que cuentan con piezas lingüísticas específicas en nuestro idioma. Por otra parte, con el fin de no hacer el repertorio demasiado largo, los marcadores que se enumeran, se remiten a veces a otra función, bajo la cual se ofrece la nómina - incompleta, como se verá - de marcadores textuales. (CASADO, 1998)

En esta misma obra, este autor proporciona una tipología de 63 funciones textuales⁶, algunas de las cuales se pueden diferenciar claramente entre sí, mientras que otras son menos

⁴**Definición de coherencia:** por coherencia se entiende la conexión de las partes en un todo con el fin de lograr un sentido determinado con el texto (CASADO, 1993).

⁵**Definición de cohesión:** La cohesión es la propiedad resultado de las funciones y formas lingüísticas destinadas a crear trabazón entre las partes del texto por alusión de unas a otras. Coseriu señala que dicho conjunto de estrategias lingüísticas no asegura la coherencia de un texto, ya que entran en juego otras habilidades más propias de la competencia expresiva y/o pragmática y/o comunicativa (CASADO, 1993).

⁶Aclaración cfr. (explicación), adición, advertencia, afirmación, aprobación, asentimiento, atenuación, (auto)corrección, causalidad, cierre discursivo, comienzo discursivo, concesividad, conclusión, condición, consecuencia, continuación, adición, contraste (cfr. concesividad, oposición), corrección (autocorrección), culminación (cfr intensificación), deducción (cfr. consecuencia), enumeración, culminación, intensificación,

claras o tienen subtipos o poseen referencias a otros tipos (por ejemplo, culminación remite en su tipología a intensificación, y deducción a consecuencia), con lo que algunos tipos se cruzan y no está descrita exactamente cuál es la relación semántica entre ellos, como ocurre en el caso de aclaración y explicación, o en el de concesividad, oposición y contraste. Además, todos estos fenómenos se conceptualizan como funciones textuales cuando remiten a otros procesos más generales sobre la comunicación que no necesariamente deben ser textuales como la expresión de la deducción, la causa o la consecuencia; otros podrían también considerarse actos de habla, como por ejemplo, la refutación, el asentimiento, la afirmación o la aprobación; y otros, finalmente, aluden más que a funciones textuales, a cuestiones estructurales como, por ejemplo, la ordenación, la continuación, el cierre o el comienzo de discurso.

3.3.3.1. Marcadores del discurso y cortesía

Por último, nos referiremos, muy brevemente, a otra parte de los estudios sobre marcadores del discurso que se ha centrado en estudiar la dimensión que su significado tiene en el nivel social de la lengua. Así por ejemplo, si pensamos en la negociación de las imágenes públicas de las identidades de los participantes que tiene lugar durante el curso de la interacción, podríamos afirmar que ningún marcador es neutro con respecto a ella y que por tanto todos ellos están implicados en mayor o menor medida en las estrategias de cortesía. Un ejemplo de ello serían los marcadores de modalización discursiva que participan en la negociación del acuerdo/desacuerdo. Atenuantes como *bueno, quizá, puede que, es posible* e intensificadores como *vale, bien, en efecto, claro que sí, por supuestísimo* juegan un papel importante en la gestión de coste y el beneficio social de la interacción (IGLESIAS, 2001).

Antes de terminar la parte teórica de este capítulo, me gustaría hacer una reflexión crítica sobre un punto de vista que domina sobre estas partículas tanto en la visión relevancista como en la Teoría de la Argumentación. Blakemore (BLAKEMORE, 1987) defiende que los enunciados poseen información representacional y computacional. Los conectores serían de este último tipo. A pesar de que caracterizan a este tipo de información no representacional con el adjetivo de computacional, ninguno de los teóricos en pragmática representan estas partículas en término lógicos. Ambos enfoques teóricos defienden que estas

consecuencia, digresión, duda, ejemplificación, énfasis, equivalencia, evidencia, exhaustividad, culminación, explicación, inclusión, ejemplificación, inferencia consecuencia, intensificación, llamada de atención, mantenimiento de atención interlocutiva, matización, negación, oposición, ordenación, enumeración, precaución, precisión, recapitulación, reformulación, refuerzo, refutación, réplica, restricción, resumen, topicalización, transición.

partículas tienen un significado procedimental y que no contribuyen a la verdad o falsedad de los enunciados ((BLAKEMORE, 1996) y (PORTOLÉS, 1998b)). Insisten en que los mensajes tienen interpretaciones potenciales y que los marcadores del discurso, desde el punto de vista de la teoría de la Relevancia, restringirían dichas potencialidades y, desde el punto de vista argumentativo, determinaría el peso argumentativo. La Gramática del Texto también ha visto cómo algunas partículas especialmente expresan relaciones entre actos de habla (DIJK, 2000). Es más, de hecho, O. Ducrot que, como sabemos, tiene una formación fuerte en lógica y en filosofía matemática (PORTOLÉS, 1998b), ha insistido explícitamente en este asunto de que los marcadores del discurso nada aportan a una semántica veritativa y no tienen cabida en la lógica; esto es lo que se ha denominado la visión no logicista del lenguaje. No obstante, una vez dicho esto, la pregunta es obligada: ¿qué se entiende realmente por computacional? ¿Qué van a hacer todas las disciplinas que dependen de una representación lógica y de un modelo computacional de la mente para dar cuenta del significado de estas palabras si no hay modelos lógicos que las representen, con las consecuencias que ello conlleva para la LC y la IA? No tengo pues más que manifestar mi desacuerdo respecto de esta cuestión. Entiendo un discurso como un encadenamiento de valores de verdad donde unos se van modificando a otros. Los marcadores del discurso influyen en la verdad de los enunciados, aunque no tengamos instrumentos de representación para dar cuenta de ellos desde la Lógica Clásica. De momento, para introducir este punto de vista que desarrollaremos convenientemente más tarde, leamos este ejemplo:

El pasado 23 de abril el Juzgado dio por concluida la investigación sobre la muerte de Mari Luz Cortés y el 8 de enero decretó el levantamiento del sumario para las partes, momento en el que se conoció que las pruebas de ADN practicadas a los pelos y tejidos hallados en algunas de las pertenencias de los acusados por este caso -Santiago del Valle, su mujer, Isabel García y su hermana, Rosa del Valle-, no los implicaban directamente en los hechos, aunque existen otras pruebas incriminatorias. El Pais Digital. Jueves, 11/6/2009.

La inferencia sobre la que en el texto se está argumentando es sobre la verdad o falsedad del enunciado implícito: Santiago del Valle et al. son los asesinos de Mari Luz. Los enunciados anteriores al marcador discursivo *aunque* argumentan en contra de esta inferencia, harían el valor falso, pero el enunciado que viene a continuación del *aunque* presenta un argumento a favor de que el valor de verdad de la inferencia sea verdadero. Creo que con este ejemplo puede quedar claro a los no especialistas en la materia cómo funcionan los marcadores del discurso dentro del significado inferido de los textos en términos de valores

de verdad.

3.4. Los marcadores discursivos, Lingüística Computacional e Inteligencia Artificial

Después de haber explicado las aportaciones teóricas más relevantes entorno al significado, función y dimensiones de análisis de esta categoría pragmática, nos gustaría introducir al lector, para finalizar este capítulo, en la aplicación que de este conocimiento se ha llevado a cabo en el mundo de la Lingüística Computacional (LC) y de la Inteligencia Artificial (IA).

Junto con las marcas superficiales de los textos, los marcadores del discurso son considerados por parte de la LC de la siguiente manera: “The most important instruments are cue words specifically employed to signal the particular relationship between text spans (STEDE y otros, 1998)”. Estas marcas resultan muy útiles para tareas de procesamiento de lenguaje natural (PLN) complejas de amplia cobertura, ya que aportan información muy rica sobre la estructura discursiva, con un bajo coste de procesamiento (ALONSO y otros, 2002b).

A continuación, enumeramos los campos de investigación en LC que se centran en estas partículas (STEDE y otros, 1998):

1. Identificación automática de la estructura discursiva de un texto. La unidad fraseológica utilizada para esta parte de la LC en inglés es *discourse structure parsing*.
2. Reconocimiento y generación de marcadores del discurso.
3. La interrelación entre la estructura del discurso y otros niveles de la gramática.
4. El análisis y la representación formal tanto de las relaciones del discurso como de sus piezas léxicas.
5. El estudio de las palabras que funcionan como marcadores del discurso en la oralidad.

Las aplicaciones de este conocimiento (ALONSO y otros, 2002b) a la LC son:

1. Resolución de co-referencia.
2. Gestión de diálogos.
3. Resumen automático.

4. Identificación de relaciones de relevancia y coherencia.
5. Segmentación discursiva de los textos.
6. Derivación de estructura discursiva.

En el panorama internacional, nos gustaría señalar dos investigaciones de gran relevancia en materia de marcadores del discurso. La primera de ellas, de carácter básica, es la creación de un recurso lingüístico, un diccionario de marcadores del discurso, que se emplea a la hora de señalar la presencia de una relación discursiva entre *spans* adjacentes de texto (BERGER y otros, 2002). En segundo lugar, destacaremos las investigaciones de D. Marcu. Este investigador ha diseñado dos algoritmos para derivar la estructura retórica de un texto. Por esto último se entiende la creación de un programa de ordenador que haría el análisis estructural de un texto, igual que los que hacen análisis sintáctico de las oraciones. El primer algoritmo identifica los marcadores discursivos en un texto y lo divide en oraciones; el segundo produce un árbol con la estructura retórica del texto. Estos algoritmos han sido diseñados para textos no restrictivos, esto es, para cualquier tipo de textos independientemente de su dominio contextual (MARCU, 1997).

En lo que respecta a las investigaciones en materia computacional en el ámbito hispano, destacaremos las investigaciones de Laura Alonso ((ALONSO y otros, 2002b) y (ALONSO y otros, 2002a)) a la hora de crear un lexicón computacional de marcadores y de usarlos en la identificación de estructura textual de cara al resumen automático de textos.

L. Alonso presenta para el resumen automático un marco y una metodología de trabajo para representar el discurso con técnicas de procesamiento de lenguaje natural. Por una lado, posee:

1. Un modelo de representación de la estructura del discurso
2. Un inventario de significados de relaciones discursivas ⁷
3. Un modo de relacionar las pistas superficiales de texto (como la puntuación) con los significados discursivos.

Por otro lado, los recursos con los que cuenta son:

1. Un corpus anotado (aunque no especifica en sus trabajos en qué niveles).
2. Un lexicón de marcadores discursivos.
3. Algoritmos para la segmentación del discurso con información varia (tampoco especifica cual).

⁷Para ver la tipología de marcadores usada por L. Alonso recomendamos (ALONSO y otros, 2002b).

4. Una herramienta para la extracción automática de marcadores del discurso en corpus sin ningún tipo de tratamiento que ha denominado MDs X-TRACTOR (ALONSO y otros, 2002b).

Para resumir, L. Alonso parte de pistas superficiales en el discurso como la puntuación, la estructura sintáctica superficial y los marcadores discursivos y los asocia a significados discursivos. El inventario de dichos significados discursivos es el siguiente: dimensión estructural, continuación, elaboración, dimensión semántica, revisión, causalidad, igualdad, contexto y discurso (ALONSO, 2005).

Una vez presentadas brevemente el estado de la cuestión en esta materia, pasemos ahora a comentar algunas deficiencias teóricas en los planteamientos computacionales. En este capítulo, solo trataremos los problemas derivados de la equivalencia entre relaciones de coherencia, relaciones retóricas y marcadores del discurso de la que se parte en buena parte de los trabajos de LC, y que ya está comenzando a ser criticada por investigadores de este mismo sector (JAYEZ y ROSSARI, 1998); mientras que dejaremos las críticas relacionadas con la identificación y clasificación de marcadores del discurso de manera manual y automática para los capítulos cinco y siete.

3.5. Relaciones de coherencia, relaciones retóricas y marcadores del discurso

He terminado la sección anterior aludiendo a las críticas que desde el terreno computacional se están arguyendo en torno a la cuestión que se rotula en este apartado. De nuevo, no deja de ser bastante sintomático de estas carencias de comunicación entre Lingüística Computacional y Lingüística el hecho de que un sector considerable de la Lingüística en los últimos años haya comenzado a manifestar enérgicamente que los marcadores del discurso pueden considerarse condición necesaria pero no suficiente para la nociones de coherencia y cohesión de los textos, mientras que desde la LC, ignorando estas afirmaciones y los argumentos que la apoyan, los marcadores del discurso sean considerados como índices que señalan las relaciones de coherencia en los textos. En bastantes estudios de LC se pueden leer afirmaciones de este tipo:

There is widespread agreement that the coherence of discourse can be characterized by certain discourse relations that can hold between adjacent spans of text.(STEDE y otros, 1998)

Así pues en la oración: *Soy un oso porque tengo canas*, el marcador del discurso *porque* hace que estos dos enunciados sea coherente, según estos planteamientos. Estos son los argumentos que han señalado C. Velardo, C. Fuentes, J. Portolés y M. Zorraquino entre otros⁸. Todos ellos muy lejos de la LC. Sin embargo, dentro de este campo, ya hemos mencionado que también ha habido voces que han estudiado la relación entre estos tres sintagmas que titulan este apartado. Una de ellas ha sido la investigadora de origen español pero afincada en EEUU desde hace bastantes años Maite Taboada, y muy adscrita en su investigación a D. Marcu. Maite Taboada en su estudio (TABOADA, 2006) discute la idoneidad de dicho planteamiento partiendo del análisis en términos de relaciones retóricas de dos corpus, uno de conversaciones y otro de artículos de periódicos. Dicha autora concluyó que entre el 60% y el 70% de las relaciones discursivas no estaban señaladas (TABOADA, 2006), lo que equivale a decir que el hablante no las ha codificado lingüísticamente; además, también concluye que hay unas relaciones que muestran más tendencia a codificarse lingüísticamente que otras, y señala los estudios de D. Marcu como uno de los sistemas que automáticamente es capaz de detectar relaciones de coherencia sin apoyarse en los marcadores del discurso.

3.6. Final de la parte teórica

Hemos expuesto brevemente en este capítulo las características definitorias de los marcadores discursivos a nivel teórico, y hemos hecho algunas puntualizaciones críticas que luego tendrán su correspondiente reflejo en nuestro sistema de anotación y el anotador semi-automático de marcadores del discurso.

Con este capítulo pues abandonamos la primera parte de la tesis, dedicada a la exposición de los conocimientos teóricos más importantes que se han elaborado en este campo de conocimientos. Antes de dar paso a la segunda parte, el módulo descriptivo, me gustaría que el lector se hubiera quedado con la impresión general de que este nivel de interpretación del significado de los enunciados está regulado por normas que solo desde hace unas décadas los investigadores se han dedicado a explicitar. Desde un punto de vista mental, el significado final de un enunciado es el resultado de una inferencia, una deducción entre lo que el enunciado significa y la fuente de conocimientos que albergamos. Así, si recibiéramos una llamada a nuestro teléfono fijo en medio de la noche, cuando ya estamos dormidos, y

⁸En palabras de J. Portolés: la cohesión no es un fin en el uso de los conectores, sino una consecuencia. Los hablantes no los emplean para lograr un texto coherente, los utilizan para que el oyente obtenga las inferencias pertinentes de los enunciados. El texto no es una unidad de la gramática como la sílaba (PORTOLÉS, 1993). Igualmente, E. Montolío señala que la cohesión no proviene de estas partículas sino de la semántica de los enunciados como también han subrayado autores como Breesler, Beaugrande, Yule y Brown (MONTOLÍO, 1998).

al descolgar, una voz seria y profunda nos dijera: *Sé lo que ocultas*, cada uno inferiría una cosa diferente, esto es, haría esa información relevante para su sistema cognitivo.

Con respecto a las reglas, me gustaría enunciar las ideas de Robin Lakoff al respecto. R. Lakoff señala que, al igual que existe una aceptabilidad de las reglas en sintáctico, y al igual que también existe una aceptabilidad del punto de vista semántico, por ejemplo, las reglas que nos hacen no producir *El oso hormiguero se dispersó*, igualmente también existen unas reglas que propician la aceptabilidad desde un punto de vista pragmático (LAKOFF, 1998). No obstante, si estas reglas existen, ¿por qué todavía no hemos conseguido sistematizar este conocimiento? Desde mi punto de vista, la razón de ello está en la metodología. De los datos abstraemos conclusiones, estas conclusiones pueden ser muy ambiguas porque todavía no hemos especificado las normas y porque siempre esas conclusiones deben estar restringidas a los contextos en los que se trabaja. Veamos un ejemplo de R. Lakoff sobre la ambigüedad pragmática de los mensajes:

1. Por favor, ¿puedes cerrar la ventana?

De este enunciado se pueden deducir dos contextos y por tanto diferentes estrategias de cortesía: en la primer interpretación, yo, que soy el hablante, tengo una relación de poder simétrica con mi oyente y por tanto no le puedo obligar a cerrar la ventana, con lo que se lo pido como un favor; en la segunda interpretación, yo, que soy el hablante, tengo una relación de poder asimétrica con mi interlocutor, y en dicha relación yo ostento la posición superior, lo que ocurre es que intento no marcar dicha relación de poder y pido como un favor lo que en principio perfectamente podría ser una orden.

Por tanto, si bien la forma lingüística codifica la norma social, las formas lingüísticas pueden estar codificando diferentes estrategias. Cualquier sistema de inteligencia artificial hombre-máquina debe en su fase de generación codificar expresiones verbales de acuerdo a normas sociales; y en su etapa de reconocimiento describir la interpretación de las formas lingüísticas en función de los contextos en los que dichas formas se están emitiendo; dentro de la descripción del contexto, como ya hemos señalado, la descripción del orden social y de cómo este se transparenta en la relación de poder entre los participantes debe ser un criterio de obligada consideración.

Y por último, y para terminar esta parte, me gustaría remarcar una idea que poco o nada se toma en cuenta en los estudios de LC y que, a mi modo de ver, es uno de los obstáculos más importantes para la Lingüística Computacional, y es que los marcadores discursivos aluden no tanto al enunciado al que siguen, sino a una de las implicaturas que

dicho enunciado despierta. Así por ejemplo, en la frase *Es pobre pero honrado*, la partícula discursiva *pero* está contraargumentando al topoi que relaciona pobreza con delincuencia y no el enunciado *Es pobre* en sí mismo. Creo que este es un buen ejemplo que ilustra que la comunicación es inferencial y que, como ya hemos señalado, los enunciados son índices de los pensamientos.

A continuación, con el siguiente capítulo, comienza lo que en la introducción denominamos el módulo descriptivo de este trabajo de tesis, el cual ocupa dos capítulos. En el primero de ellos explicaremos detalladamente el modelo de anotación pragmática, mientras que en el segundo, mostraremos nuestra clasificación semántico-pragmática de los marcadores discursivos encontrados en el corpus C-ORAL-ROM.

Parte II

De la descripción

Capítulo 4

Pragmatext, un modelo de anotación pragmática para corpus

4.1. Introducción

Presentamos en este capítulo los principales fenómenos lingüísticos que vamos a tratar de reconocer y etiquetar en el corpus. Dichos fenómenos son los protagonistas de nuestro modelo de anotación. Como señala H. López Morales (LÓPEZ, 1994), “los modelos son construcciones teóricas hipotéticas, son teorías provisionales preparadas para su verificación empírica”. La validación de este modelo se llevará a cabo en el capítulo siguiente, cuando clasifiquemos semánticamente una a una las unidades con información pragmática del corpus C-Oral-Rom.

El capítulo está dividido en dos grandes apartados. En el primero, se explica la conceptualización general del modelo y su implementación a lenguaje XML; en el segundo apartado, explicamos más detenidamente cada uno de los fenómenos que vamos a etiquetar.

4.2. Concepción teórica del modelo de anotación pragmática

Como ya se apuntó en el capítulo primero, la motivación principal que sostiene este trabajo de tesis es la reflexión sobre el posible tratamiento que se le puede dar desde una perspectiva computacional a esta *nueva* consideración cognitivo-social y pragmática del lenguaje.

En español, los fenómenos que vamos a tratar se manifiestan a través de recursos lingüísticos que afectan a todos los niveles de la gramática. Este hecho nos obliga a seguir un camino diferente al que tradicionalmente ha seguido la Lingüística Computacional. En

lugar de remontarnos al nivel semántico a partir de las categorías gramaticales, hemos ido de los fenómenos a las formas lingüísticas, independientemente de la categoría gramatical a la que pertenezcan. Esperamos ilustrar con suficiente claridad tanto nuestra concepción teórica como nuestra metodología de análisis con el siguiente cuadro:

COGNICIÓN EN INTERACCIÓN	PRAGMÁTICA		LINGÜÍSTICA	
	Fenómenos pragmáticos	Categorías pragmáticas	Categorías gramaticales	Formas lingüísticas
Estrategias de cognición y comunicación		Coargumentación, concesión, hipótesis, reformulación, etc.	Conjunciones, Interjecciones, Adverbios oracionales, SSPP, etc.	y, además, aunque, si, o sea, etc.
¿Cómo razonamos?	Relaciones Argumentativas			
¿Cómo moldeamos la percepción en la interacción social?	Modalización discursiva	Intensificación, Atenuación e Interacción	Idem	segurísimo, bueno, ¿sabes?
¿Cómo sabemos lo que sabemos?	Evidencialidad	Vista, otros sentidos, inferencia, otra persona, fuente oral	Idem	generalmente, por lo visto, quien a buen árbol se arrima...
¿Cómo expresamos las emociones verbalmente?	Lenguaje de las emociones	positivas, negativas, sorpresa	idem	por fortuna, por desgracia, qué fuerte
¿Cómo convencionalizamos la percepción?	Metáfora	Dominio Origen, Dominio Destino	Idem	por un lado
¿Cuáles son nuestras intenciones?	Actos de habla	Saludos, etc.	Idem	perdón, ánimo, vengas, etc.
¿Cómo nos referimos a otras partes del discurso y al interlocutor?	Deixis	Textual y Social	Idem	anteriormente, tío, pibe, macho

Tabla 4.1: Concepción teórica del modelo de anotación pragmática para corpus Pragmatext

Como vemos, lo que queremos ejemplificar en la tabla es cómo un fenómeno que tiene un origen social y cognitivo queda modelado por el conocimiento teórico procedente de la Pragmática y de la Semántica, y se termina finalmente materializando en formas lingüísticas concretas pertenecientes a diferentes categorías establecidas por la Gramática Tradicional.

4.3. Descripción de los fenómenos semántico-pragmáticos de PRAGMATEXT

A continuación, presentamos una breve descripción de los siguientes fenómenos:

1. Lenguaje emocional
2. Relaciones discursivas
3. Modalización lingüística
4. Evidencialidad
5. Deixis
6. Actos de habla
7. Metáforas y unidades fraseológicas

La exposición de cada fenómeno va a seguir siempre la misma estructura discursiva:

1. Descripción del fenómeno.
2. Sistematización del mismo en el modelo de anotación, Pragmatext.
3. Ejemplos de categorías gramaticales que expresan este fenómeno.
4. Ejemplos de formas lingüísticas concretas.
5. Formalización de la anotación del fenómeno a lenguaje XML.

4.3.1. La codificación verbal de las emociones

La palabra *emoción* viene del latín *emotio*, que significa *el impulso que induce la acción*. En buena parte de la tradición filosófica, las dimensiones de la experiencia cognitiva intelectual o racional fueron consideradas separadas y superiores a la emocional o sentimental. Paralelamente se estableció que el conocimiento y la razón pertenecen al hombre mientras que las emociones y los sentimientos a la mujer. Actualmente, hay un acuerdo en

todas las disciplinas para rechazar esta falsa dicotomía (emoción vs. razón). La Psicología, la Neurolingüística y la IA están muy interesadas en saber más cosas de las emociones, ya que éstas juegan un papel fundamental en la toma de decisiones en la vida cotidiana, en el almacenamiento y evocación de memorias (MORA, 2001) y en la comunicación entre mundo externo e interno (DALGLEISH y POWER, 1999). Desde la IA, los investigadores que se han interesado en las emociones proceden de las áreas de los sistemas de diálogo hombre-máquina y del diseño de sistemas expertos. Ambos coinciden en la necesidad de tomar en cuenta las emociones como un dato para limitar las infinitas alternativas posibles de la elección racional (MAFFIA, 2005).

En este trabajo, estamos interesados en la relación entre lenguaje y emoción. **La Psicología** reconoce que tanto el lenguaje como las emociones son los dos sistemas de comunicación del ser humano por excelencia; sin embargo, existen en su tradición pocos estudios que intenten analizar cómo estos dos sistemas de comunicación interactúan entre sí (DALGLEISH y POWER, 1999). Desde La Psicología de las Emociones, estas han sido estudiadas a través de las expresiones faciales y a través del lenguaje gestual (NISSAN, 1997).

De la tradición de **estudios filosóficos y psicológicos**, lo que más nos interesa es la tipología de emociones básicas que se ha esbozado. Entre las fuentes consultadas, los estudios coinciden en señalar las siguientes emociones: la alegría, la ira, la tristeza y el miedo (DALGLEISH y POWER, 1999) (GÜELL, 2006). Algunos incluyen además el pánico y el disgusto (DALGLEISH y POWER, 1999).

Por otra parte, tampoco debemos dejar de citar las reflexiones en torno a la expresión lingüística de las emociones que se han hecho en **la Sociología o la Antropología**, y que, en definitiva, inciden en la idea de que el lenguaje convencionaliza la expresión de las emociones, y las normas culturales y sociales se encargan de regular la expresión de las mismas (ISRAEL, 2006) ¹.

Desde la **Lingüística**, las emociones tradicionalmente se han abordado a través del estudio del paralenguaje, la prosodia y la interjección. Desgraciadamente, nuestro corpus no goza de un rico sistema de marcas prosódicas, y solo diferencia unidades tonales y enunciados, como ya hemos señalado; no marca, por tanto, entonación ascendente, descendente o sostenida, ni tampoco las palabras que se pronuncian con un énfasis especial; de hecho, la entonación exclamativa no se marca, con lo que no podremos abordar este tema en este nivel del lenguaje a partir del cual se podría estudiar cómo la prosodia codifica estados emocionales. Por contrapartida, donde sí podemos analizar la expresión de las emociones

¹En relación con el papel de las culturas en la regulación de las emociones, Punset (PUNSET, 2006) define la Biblia y el Corán como propuestas pedagógicas para controlar emociones.

es a través de la interjección y otras formas lexicalizadas.

Desafortunadamente, buena parte de los estudios consultados sobre la interjección se ocupan en gran medida del estatus gramatical de dicha categoría. Debate que no vamos a tratar en este trabajo, ya que no forma parte de los objetivos de la investigación. No obstante, la Pragmática ha dado un nuevo impulso a la explicación semántica de la interjección. Entre las aportaciones más importantes de estos nuevos marcos de análisis, está la concepción de estas partículas, primero, como partículas de modalidad, esto es, formas de expresión de la actitud del hablante respecto de lo enunciado, y segundo, como agentes restrictores de las inferencias contextuales a las que conduce un enunciado ² (CUETO y LÓPEZ, 2003). Veamos el siguiente ejemplo:

1. Zapatero ha ganado otra vez las elecciones.

a) Vaya por dios

b) ¡Hurra!

c) ¡Qué fuerte!

Como se puede observar, no solamente es importante el contenido proposicional de un enunciado, del cual las interjecciones no nos dicen absolutamente nada, sino también el estado emocional que la verdad o falsedad del mismo produce en los participantes de la interacción: disgusto en (a), alegría en (b) o sorpresa en (c).

Sirva este ejemplo para constatar, primero, la importancia que tienen esta clase de palabras en la comunicación, y, segundo, la necesidad de profundizar en su significado, distribución, y dimensiones cognitiva, social y cultural.

Lamentablemente, a pesar de los avances señalados en esta materia, no encontramos estudios globales sobre cuántas interjecciones hay en español, y qué significa exactamente cada una de ellas. Las tipologías encontradas se centran en el contenido que dichas partículas codifican, el cual no es solo de naturaleza emocional, sino también de naturaleza deíctica (*¡cuidado!*) (WIKINS, 1992), interactiva o fática (*oye*) (AMEKA, 1992) e ilocutiva, esto es, codifican acciones (*¡vamos allá!*) (ALONSO, 1999). Sin embargo, no encontramos estudios sistemáticos sobre las interjecciones, ya que la mayoría de los autores consultados terminan alegando que su significado último viene condicionado por el contexto (TORRES, 2000).

²Según I. Cuetos, (CUETO y LÓPEZ, 2003) son restrictoras de explicaturas de alto nivel dentro del marco la teoría de la Relevancia.

En **Inteligencia Artificial**, ya existen proyectos de investigación orientados al reconocimiento de los estados emocionales de los hablantes, especialmente de los estados negativos, ya que el reconocimiento de los mismos puede ser una pista sobre el éxito o fracaso de la comunicación, de modo que el sistema pueda desarrollar estrategias de redireccionamiento en sus acciones (FISCHER, 1999) (KOSTOULAS y FAKOTAKIS, 2008).

Así pues, para resumir lo dicho hasta ahora, existen bastantes parcelas de conocimiento interesadas en el estudio de las emociones, por su trascendencia en el razonamiento y en la comunicación; sin embargo, hay problemas para analizarlas desde un punto de vista empírico. La Lingüística hasta ahora ha profundizado un poco más en el significado de la categoría expresadora de emociones por excelencia, que es la interjección, aunque dicha categoría no solo exprese emociones como ya hemos señalado; no obstante, la ambigüedad contextual en la interpretación de su significado y el hecho de que no tengan un referente claro, sino más bien tipos de situaciones asociadas a su aparición (CUETO y LÓPEZ, 2003), impide una descripción clara y concisa de su semántica que pueda ser útil a los no lingüistas como, por ejemplo, los ingenieros o los profesores de lenguas.

Con nuestra propuesta, llegamos a un compromiso entre la búsqueda de simplicidad de las exigencias del punto de vista computacional y el relativismo basado en la riqueza de significados o valores que podemos diferenciar desde un punto de vista lingüístico. Vamos a anotar las categorías que expresen una evaluación de la realidad en términos positivos o negativos o de sorpresa: los adverbios *lamentablemente* y *afortunadamente* y el sintagma exclamativo *pero qué me estás contando* ejemplificarían los tres tipos.

Esta medida simplificadora no impide que la clasificación de las interjecciones y expresiones emocionales siga siendo ambigua. Por ejemplo, la interjección *¡joder!* puede codificar una emoción negativa ante un estímulo, pero, en algunos casos, también puede implicar euforia o sorpresa. En nuestro etiquetado, puede tener por tanto los tres valores y se desambiguará manualmente cuál es cuál según el contexto de realización, ya que no nos podemos ayudar de la prosodia, la cual sería en este caso de una gran ayuda. De todas formas, ya veremos en el capítulo de la clasificación semántica de las unidades pragmáticas del corpus hasta qué punto se reduce así el problema de la ambigüedad de valores en el significado de estas palabras.

Las **categorías gramaticales** encargadas de llevar a cabo la expresión de las emociones localizadas en nuestro corpus han sido las siguientes: interjecciones y adverbios oracionales, construcciones exclamativas como, por ejemplo, *qué guay*, adjetivos como *chévere* y sintagmas preposicionales como *de puta madre*.

Algunos **ejemplos de formas lingüísticas** en las que se materializan esta operación son: *afortunadamente*, *desgraciadamente*, *¡yeah!*, *ole*, *bravo*, *guay*, etc.

En la formalización del sistema de anotación a lenguaje XML, hemos abierto un campo para las emociones que se ha codificado con el atributo ED y que puede adoptar tres valores: positivo, negativo³ o sorpresa⁴. A continuación presentamos un ejemplo de una forma léxica de C-Oral-Rom con significado emocional en formato XML:

```
<DM FU="Loc" MET="No" DR="No" ED="Surprise" MOD="No"  
EVI="No" SA="No" DEX="No">madre del amor hermoso!</DM>
```

4.3.2. Las operaciones argumentativas

Argumentar es la estrategia por medio de la cual un hablante, expresándose en una lengua natural, llega a sacar conclusiones válidas (Stati 1990, en Fuentes (FUENTES y ALCAIDE, 2002)). Para Ducrot y Anscombre (ANSCOMBRE y DUCROT, 1983), **“argumentar es dar a A una razón para creer C”** (FUENTES y ALCAIDE, 2002). En este trabajo, estamos interesados en la argumentación como un proceso cognitivo de organización del razonamiento (GILLE, 2001). Las operaciones argumentativas serán aquellas que cumplan la función de llevar a cabo esta labor. Ya hemos mencionado en el capítulo uno y tres cómo los marcadores discursivos ofrecen una cantidad de matices semánticos en la expresión discursiva del razonamiento que las conectivas lógicas no expresan. Así por ejemplo, en la Lógica Clásica, el marcador contraargumentativo *pero* pertenecería a la misma conectiva formal que la conjunción copulativa y (ESCANDELL, 1996).

La tipología de operaciones argumentativas que se expondrá a continuación tiene una base teórica basada en la Gramática del Texto y en las corrientes pragmáticas citadas en el capítulo anterior. Algunos de los tipos remiten a categorías cognitivas que tienen que ver con operaciones de pensamiento reconocidas por la Lógica, tales como la deducción y la inducción. Otros, en cambio, como la planificación discursiva o la reformulación, son propios de la actividad formulativa del discurso.

Estos son los tipos de operaciones argumentativas que hemos distinguido a partir de la anotación de los marcadores del C-ORAL-ROM:

³Como el etiquetado se ha elaborado en lengua inglesa, las etiquetas en realidad son: Positive Evaluation, Negative Evaluation y Surprise.

⁴J. A. Marina (MARINA, 1993) incluye la sorpresa dentro de las emociones básicas y la define como el sentimiento producido por una novedad. Los psicólogos dicen que la sorpresa sólo puede darse si esperamos algo, un sentimiento producido por la inadecuación de lo percibido con lo esperado. En Pragmatext, la sorpresa será una categoría pragmática.

1. Generalización y Concreción.
2. Coargumentación, Contraargumentación, Concesión y Opcionalidad.
3. Hipótesis-Condición.
4. Causa y Consecuencia.
5. Reformulación.
6. Topicalización.
7. Síntesis y Conclusión.
8. Planificación del Discurso.
9. Digresión.
10. Finalización discursiva: Definida, Indefinida y de Transición.
11. Cambio de tópico.
12. Escalaridad.
13. Marco.

4.3.2.1. Generalización y Concreción

Las operaciones de generalización y concreción ⁵ discursiva responderían en el nivel cognitivo con el pensamiento inductivo y deductivo, dos de los razonamientos claves en las computaciones que el cerebro realiza durante el procesamiento de información. Los procesos de metáfora y los refranes son ejemplos del uso cognitivo de estos dos tipos de razonamiento, como se explicará más detalladamente un poco más abajo.

Las **categorías gramaticales** que hemos localizado en el corpus encargadas de llevar a cabo estas funciones han sido las siguientes: sintagmas preposicionales, adverbios oracionales y conjunciones.

Algunos **ejemplos de formas lingüísticas** en las que se materializan estas operaciones discursivas de origen cognitivo son: *en general, en particular, generalmente, por regla general, en teoría, en concreto, concretamente, en particular, etc.*

A continuación presentamos dos **ejemplos en lenguaje XML**, uno de cada tipo:

1. <DM FU="Loc" MET="No" DR="Concretion"
ED="No" MOD="No" EVI="Inference" SA="No" DEX="No">en concreto</MD>

⁵Generalization y Concretness.

2. <DM FU="Loc" MET="No" DR="Generalization"
 ED="No" MOD="No" EVI="Inference"
 SA="No" DEX="No">por lo general</MD>

4.3.2.2. Coargumentación, Contraargumentación, Concesión y Opcionalidad

La **coargumentación y la contraargumentación**⁶ son dos tipos de operaciones que encuentran su trasfondo teórico en la Teoría de la Argumentación. Según dicha teoría, los enunciados tienen un peso argumentativo a la hora de conducir a determinadas verdades o conclusiones sobre la realidad. Los enunciados que aumentan el peso argumentativo de otros estarán orientados, mientras que aquellos que intenten contrarrestarlo y cambiar la dirección de las conclusiones que interlocutor debiera sacar del primer enunciado serán contraargumentativos.

Dentro de la coargumentación, hemos diferenciado tres tipos diferentes con el fin de dar cuenta de la enumeración de argumentos dentro del discurso: coargumentación1 (*en primer lugar, primero*), coargumentación2 (*en segundo lugar, tercer lugar, y, también, además, después, etc.*), y coargumentación3 (*en último lugar*). De esta manera, grosso modo, se representarían los argumentos que aparecen al inicio, durante y en último lugar en un conjunto de argumentos o enunciados dedicados a defender un valor de verdad.

Muchos de estos marcadores tienen contenido metafórico e indican una conceptualización del discurso en partes, en lados o pilas: *por un lado, por otra parte, en primer lugar, etc.* Como es de esperar, dichas conceptualizaciones del discurso pueden variar de una lengua a otra. Por ejemplo, en inglés, la metáfora que concibe los argumentos en lados también existe (*on one side*), pero también se utiliza el cuerpo como dominio de partida; en concreto, la mano (*on the other hand*). Pensamos que es importante reflejar el contenido metafórico de estas palabras, a fin de favorecer la comparabilidad entre las lenguas. Es por eso que lo incluiremos en el etiquetado de estas partículas a través de un rasgo que luego será en lenguaje XML un atributo.

La concesión es una operación a mi modo de ver polifónica. Muestra la consideración de otros posibles argumentos (procedentes de otra voz discursiva) contra la posición defendida; a diferencia de la contraargumentación, los argumentos presentados después de un marcador concesivo presentan una carga argumentativa menor, insuficiente para que conduzca a otras conclusiones. Por ejemplo, en el enunciado *Aunque los inmigrantes puedan votar, nunca tendrán los mismos derechos*, el primer enunciado *Los inmigrantes puedan votar* representa un argumento que pertenece a una voz discursiva que defiende el discurso

⁶Denominación en inglés de los tipos establecidos: Coargumentation1, Coargumentation2, Coargumentation3, Contraargumentation, Concessiveness, Anaphoric-Concessiveness-Cataphoric-Conclusion, Option.

de *Todos somos iguales*; diferente al discurso al que alude el enunciado *Nunca tendrán los mismos derechos*. El primer enunciado, cuando se conecta con un concesivo, se presenta como un argumento digno de tomar en consideración, una argumento basado en otra voz, pero que finalmente es rebatible por el argumento que viene a continuación *Nunca tendrán los mismos derechos*.

Categorías gramaticales encargadas de llevar a cabo estas funciones son: conjunciones y sintagmas preposicionales.

Algunos **ejemplos de formas lingüísticas** en las que se materializan esta operación son: *aunque*, *a pesar de ello* y *pese a que*.

A continuación presentamos un **ejemplo en lenguaje XML**:

```
<DM FU="Loc" MET="Yes" SD="Weight" TD="Concessiviness"
DR="Concessiviness" ED="No" MOD="No" EVI="No" SA="No"
DEX="No">a pesar de que</MD>
```

Un marcador discursivo del tipo *aunque* presenta los argumentos considerados pero sin peso suficiente a continuación de dicho marcador. Sería una concesividad catafórica, que hemos denominado simplemente concesividad por ser la más prototípica. No obstante, hemos tenido que hacer un nuevo tipo de concesividad, **concesividad anafórica y conclusión catafórica**, para clasificar semánticamente a todos aquellos marcadores discursivos que presentan todos los enunciados anteriormente señalados como insuficientes para la conclusión que ellos introducen. Por ejemplo:

Los iraníes están manifestándose en las calles contra el pucherazo. El consejo de guardianes ha juzgado fraudulentos tres millones de votos. De todas formas, las elecciones no se repetirán.

Los dos argumentos anteriores, orientados al valor de verdad *Las elecciones han sido un fraude*, no son suficientes para concluir, ya que el marcador discursivo atenúa dicha fuerza argumentativa y concluye el valor de verdad contrario.

Finalmente, hemos definido un tipo nuevo para aquellos marcadores que indiquen **opcionalidad**, como por ejemplo, la conjunción *o* u *o bien*.

4.3.2.3. Hipótesis y condición

Tanto la hipótesis ⁷ como la condición han sido bastante estudiadas tanto en Lingüística como en Lógica a partir de la partícula *si*. De hecho, buena parte de los estudios consultados están dedicados a comparar el *si* de las lenguas naturales con el *si* lógico. Sin embargo, al igual que en el caso de las conjunciones, los estudios en Pragmática tampoco han dado muchas esperanzas a la hora de establecer una posible equivalencia semántica.

En el mundo hispánico, la investigadora que más ha profundizado en el estudio de esta partícula ha sido Estrella Montolío (MONTOLÍO, 1993). Los contenidos que citamos a continuación corresponden a las publicaciones posteriores a su trabajo de tesis sobre las oraciones condicionales.

El marcador discursivo *si* expresa varios significados, entre ellos, los más importantes son la condición y la hipótesis. Sin embargo, tenemos que atender a usos no muy tratados ni desde la perspectiva de la Lógica Clásica ni desde la Gramática tradicional como la presencia del *si* en estructuras condicionales unimembres del tipo *Pero si yo no he sido*, donde parece que estos significados se explican mejor a la luz de otras teorías, como por ejemplo, la argumentativa, que dentro de los marcos establecidos por la Gramática Tradicional. Además, Montolío (MONTOLÍO, 1993) destaca la relación que existe entre atenuación e hipótesis, explicándonos cómo los hablantes eligen plantear una aseveración a modo de hipótesis, bien con el fin de mantener las relaciones sociales con su interlocutor, bien con el fin de no aseverar algo de lo que no se está del todo seguro.

En Pragmatext, hay un rasgo exclusivo para la hipótesis y otro para la atenuación; en nuestro modelo los dos rasgos siempre estarán activados para el caso de la partícula *si*. La razón de ello es porque no tenemos ninguna manera de saber cuándo el hablante está planteando la hipótesis porque la ha concebido ordinariamente así o bien porque decide, elige conscientemente esa forma, en lugar de una afirmación, por motivos sociales. Lo que está detrás de esta elección, como veremos cuando lleguemos a la modalización, es el grado de responsabilidad que el hablante decide asumir frente a la fuerza aseverativa de su enunciado. De igual manera, no hemos distinguido entre aquellos usos que entrarían dentro del dominio de la hipótesis, y aquellos usos que indican condición, ya que esto nos obligaría al etiquetado manual de todas las ocurrencias de esta palabra, y esta tesis no es un estudio de la partícula *si*, sino que es la creación de un recurso para que en el futuro puedan estudiarse de una manera más empírica a partir de un corpus oral dichos significados hipotéticos y condicionales.

Las **categorías gramaticales** encargadas de llevar a cabo estas funciones son: conjun-

⁷Hypothesis

ciones, sintagmas preposicionales y adverbios.

Algunos **ejemplos de formas lingüísticas** en las que se materializan esta operación son: *si, acaso, a ver si, por si acaso, etc.*

A continuación presentamos un **ejemplo en lenguaje XML**:

```
<MD FU="No" MET="No" DR="HypothesisCondition" ED="No"
MOD="Hedge" EVI="No" SA="No" DEX="No">si</MD>
```

4.3.2.4. Causa y consecuencia

La causalidad ⁸ está considerada como una categoría de pensamiento. Los estudios clásicos de Piaget sobre el funcionamiento de la causalidad durante el desarrollo cognitivo del niño son una buena muestra de la atención científica que ha recibido este fenómeno. Si tal y como es sobradamente entendido, el pensamiento racional se basa en la causalidad, no es de extrañar por tanto que las lenguas reserven un buen número de estrategias lingüísticas para expresar la causalidad y sus efectos.

Las categorías gramaticales encargadas de llevar a cabo estas funciones son: conjunciones, sintagmas nominales, preposicionales y adjetivales, si es que se entiende que las palabras *dado* y *puesto* en las construcciones *dado que* y *puesto que* pueden ser consideradas adjetivos.

Algunos ejemplos de **formas lingüísticas** en las que se materializan esta operación son: *porque, gracias a, dado que* y *puesto que*. Para la consecuencia: *en consecuencia, consecuentemente, como consecuencia* y *por lo tanto*.

A continuación presentamos dos **ejemplos en lenguaje XML**, uno de cada tipo:

1. <MD FU="Col" MET="No" DR="Cause" ED="No"
MOD="No" EVI="No" SA="No" DEX="No">debido a que</MD>
2. <MD FU="Loc" MET="No" DR="Consequence"
ED="No" MOD="No" EVI="No" SA="No" DEX="No">en consecuencia</MD>

La doble direccionalidad semántica de algunas partículas de esta clase nos ha obligado a explicitarla en dos subtipos: primero, los causales catafóricos; y, segundo, los que son al mismo tiempo causales anafóricos y consecuenciales catafóricos. Ejemplo del primero es el marcador *resulta que*; ejemplo del segundo: *de ahí que*.

⁸Cause, Consequence, Anaphoric-Cause-Cataphoric-Consequence, Cataphoric-Cause.

4.3.2.5. Reformulación

La reformulación⁹ es una operación discursiva mediante la cual el hablante explicita inferencias haciendo que un enunciado sea una paráfrasis del anterior (GÜLICH y KOTSCHI, 1983). La bibliografía ha distinguido diferentes tipos de operaciones de reformulación, algunas de las cuales conviven dentro de un mismo marcador, como es el caso de *o sea*, por ejemplo (FÉLIX-BRASDEFER, 2006). En Pragmatext, solo vamos a marcar si una palabra sirve para reformular. No vamos a abrir subtipos dentro de la reformulación, ya que antes de diferenciar los diversos tipos de reformulación diferenciados por la descripción que la Pragmática ha dado del fenómeno, primero debemos plantearnos, a nivel computacional, el reconocimiento de esta operación y sus consecuencias en términos del procesamiento y comprensión final de los enunciados que participan en una reformulación discursiva.

Las **categorías gramaticales** encargadas de llevar a cabo estas funciones son: adverbios o sintagmas adverbiales y sintagmas preposicionales.

Algunos **ejemplos de formas lingüísticas** en las que se materializan esta operación son: *o sea, esto es, mejor dicho, a saber, etc.*

A continuación presentamos un **ejemplo en lenguaje XML**:

```
<MD FU="Loc" MET="No" DR="Reformulation" ED="No" MOD="No" EVI="No"
SA="No" DEX="No">o sea</MD>
```

4.3.2.6. Topicalización

Pertenecen a esta clase las formas lingüísticas¹⁰ que convierten en tópico una agrupación o chúnker de información determinada.

Las **categorías gramaticales** encargadas de llevar a cabo estas funciones son: sintagmas nominales y preposicionales.

Algunos **ejemplos de formas lingüísticas** en las que se materializan esta operación son: *respecto a, con referencia a, respecto de, etc.*

A continuación presentamos un **ejemplo en lenguaje XML**:

```
<MD FU="Loc" MET="No" DR="Topicalization" ED="No" MOD="No" EVI="No"
SA="No" DEX="No">repecto a</MD>
```

⁹Reformulation

¹⁰Topicalization

4.3.2.7. Síntesis y Conclusión

Ante las dificultades de la bibliografía para distinguir nítidamente entre una y otra operación discursiva, hemos decidido hacer una gran clase¹¹ donde se incluyan todos los marcadores que sirvan para concluir el discurso, bien a modo de resumen, bien a modo de conclusión.

Las **categorías gramaticales** encargadas de llevar a cabo estas funciones son: nombres, adverbios y sintagmas preposicionales.

Algunos **ejemplos de formas lingüísticas** en las que se materializa esta operación son: *en suma*, *en conclusión*, *en dos palabras* y *en resumidas cuentas*.

A continuación presentamos un **ejemplo en lenguaje XML**:

```
<MD FU="Col" MET="No" DR="Conclusion_Synthesis" ED="No"
MOD="No" EVI="No" SA="No" DEX="No">en una palabra</MD>\\
```

Dentro de esta clase, encontramos el siguiente paisaje discursivo:

1. Palabras que se destacan por tener un significado puramente de resumen, como por ejemplo, *en una palabra o brevemente*.
2. Palabras que tienen un significado de conclusión claro, sin la connotación del resumen, como por ejemplo, *en conclusión*.
3. Palabras que se pueden entender como una mezcla de ambas operaciones semánticas como *en total*.
4. Palabras que comparten parte de su semántica con la operación de resumen, pero también de focalización, como por ejemplo, *principalmente*, o *básicamente*.
5. Y, por último, palabras como *así que todo eso*, que tienen una parte de consecuencia.

Véase con estos ejemplos que estas palabras tienen una semántica para nosotros compleja de describir, ya que las podemos encajar en diferentes clases; es en los contextos donde se pone de relevancia, repunta, una operación semántica por encima de otras que también están, por decirlo así, en su código genético.

¹¹Conclusion-Synthesis.

4.3.2.8. La planificación del discurso

Entran dentro de este tipo ¹² todas las formas lingüísticas que se ocupan de sostener y planificar la intervención verbal.

Las categorías gramaticales encargadas de llevar a cabo estas funciones son: signos paralingüísticos, sintagmas verbales y preposicionales.

Algunos ejemplos de **formas lingüísticas** en las que se materializan esta operación son: *&ah* y *&eh*.

A continuación presentamos un **ejemplo en formato XML**:

```
<MD FU="No" MET="No" DR="Dicourse_Production"
ED="No" MOD="No" EVI="No" SA="No" DEX="No">ah</MD>
```

4.3.2.9. Digresión

La digresión ¹³ es la operación discursiva mediante la cual se presenta una secuencia o enunciado que se aparta del hilo discursivo principal y que ha sido sugerida por algo que se acaba de decir; suele ser espontánea o no planificada. En (ACÍN, 1999) se caracteriza como una operación que abre una puerta temática de carácter breve, pero debemos recordar que, tal y como funciona la conversación espontánea, la digresión a veces puede abrir un nuevo hilo temático y que probablemente no se vuelva al anterior.

Categorías gramaticales: sintagmas preposicionales.

Ejemplos de formas lingüísticas: *por cierto*, *a propósito*.

Ejemplo en formato XML:

```
<MD FU="Col" MET="No" DR="Digression" ED="No"
MOD="No" EVI="No" SA="No" DEX="No">por cierto</MD>
```

4.3.2.10. Finalización discursiva: definida e indefinida y de transición

Esta operación del discurso ¹⁴ no descrita hasta el momento por la bibliografía se centra en la zona final de los enunciados y se preocupa de clausurar las argumentaciones (finalización definida), o bien de favorecer en la narración de los hechos la transición de unos enunciados a otros mediante partículas que contienen un buen grado de atenuación semántica, como es el caso de *y nada*, *y tal*, *pues nada*.

¹²Discourse Production.

¹³Digression

¹⁴Denominados en la DTD DefiniteEnd, IndefiniteTransitionalEnd

Categorías gramaticales: sintagmas preposicionales y construcciones atípicas para la clasificación tradicional de categorías como la palabra *pues nada*.

Ejemplos de formas lingüísticas: *y punto*, *y eso*, *y nada*, *y tal*, *y ya*, *y ya está*.

Ejemplo en formato XML:

```
<MD FU="Col" MET="No" DR="Definite_End" ED="No"
MOD="No" EVI="No" SA="No" DEX="No">y punto</MD>
```

4.3.2.11. Escalaridad

Conforman esta clase ¹⁵ los marcadores que tienen un significado escalar que ya explicamos en el capítulo dos.

Categorías gramaticales: adverbios.

Ejemplos de formas lingüísticas: *inclusive*, *incluso*.

Ejemplo en formato XML:

```
<DM FU="No" MET="No" SD="No" TD="No" DR="Scalar_Meaning" ED="No"
MOD="No" EVI="No" SA="No" DEX="No">inclusive</DM>
```

4.3.2.12. Marco

En nuestro modelo de anotación, el marco ¹⁶ es una operación de enmarcación contextual a través de la cual se expresa que la verdad de dicho enunciado solo es válida dentro de las condiciones que impone el marco.

Las **categorías gramaticales** encargadas de llevar a cabo estas funciones son: adverbios y sintagmas preposicionales.

Algunos **ejemplos de formas lingüísticas** en las que se materializan esta operación son: *desde el punto de vista de X* y *en este sentido* pero también *en tanto en cuanto*.

A continuación se presenta un **ejemplo en lenguaje XML**:

```
<MD FU="Col" SD="Sight" TD="Meaning" MET="No"
DR="Framework" ED="No" MOD="No" EVI="No"
SA="No" DEX="No">desde el punto de vista</MD>
```

¹⁵ScalarMeaning

¹⁶Framework

4.3.3. La modalización discursiva

La modalización discursiva ha sido objeto de estudio de diferentes ramas del conocimiento, desde la Filosofía hasta la Pragmática. A pesar de que la Lógica y la Semántica formal diferencia entre distintos tipos de modalidad, solo nos vamos a centrar en la epistémica, y en Pragmatext, la llamaremos modalización para aludir al proceso de moldeamiento de nuestra percepción respondiendo a intereses cognitivos y sociales.

La modalización discursiva hace referencia a la codificación del grado de certeza que el hablante muestra con la verdad de los enunciados que está profiriendo. Hemos diferenciado dos tipos de operaciones de modalización: la **intensificación** y la **atenuación**. Por la primera, entendemos la implicación explícita del hablante con lo que está diciendo. Por atenuación, en cambio, entendemos el distanciamiento del hablante, la no asunción de responsabilidad con la verdad de sus enunciados. En el capítulo primero ya aludimos a las razones por las cuales se modaliza el discurso. Una causa puede ser de origen cognitivo: el hablante no tiene certeza sobre lo que está diciendo y así se lo hace saber a su interlocutor. La otra es de origen social: puede que asumir la verdad signifique asumir una responsabilidad social, o puede que signifique entrar abiertamente en conflicto con el interlocutor.

La atenuación y la intensificación pueden afectar a más actos de habla que a la aseveración. Esta relación entre mitigación y actos de habla ha sido ya tratada por la bibliografía (CAFFI y JANNEY, 1994). En el enunciado *Me parece que no has hecho bien*, el segmento *me parece* se utiliza para atenuar un juicio, una censura o un reproche (HAVERKATE, 1994). Haverkate (HAVERKATE, 1994) destaca cómo el holandés tiene para la exhortación partículas de modalidad como *Enns eve Maar* que atenúan la fuerza del imperativo. En español, la forma *haz el favor* en *Ven aquí, haz el favor*, también sería un atenuante del imperativo.

Por otra parte, la atenuación también se materializa en otros niveles gramaticales como, por ejemplo, el morfológico con el uso del morfema *-ita* en el caso de *¿Tomamos una cervecita?*, donde lo que en realidad se está atenuando es el coste social que conlleva el acto de la oferta o invitación propuesto en el enunciado.

Por otra parte, dentro de las operaciones de modalización discursiva, nos gustaría diferenciar un tipo más: **las formas interactivas**. Es posible que muchos teóricos no se muestren de acuerdo con la inclusión de esta categoría dentro de este fenómeno. Ya Tuson (VIAGARA, 1990) señala los problemas para integrarla teóricamente dentro de la estructura de la comunicación. Las formas apelativas o fáticas como *¿sí?* *¿no?* *¿sabes?* *¿no me digas?* tradicionalmente han sido consideradas como ejemplos de la función fática del lenguaje. Según esta autora, en los enunciados con función fática, el lenguaje está orientado a informar acerca del uso que del canal de comunicación se hace en el acto comunicativo,

y la función que cumpliría sería la de asegurar la atención y la comprensión del oyente. Desde mi punto de vista, no sólo el hablante busca asegurarse de la atención y de la comprensión del mensaje, sino que también busca el acuerdo con la verdad del enunciado que se está profiriendo. Así, en el ejemplo: *Cojo las entradas para el concierto ¿no?*, es el acuerdo sobre las acciones que implican asentir a la verdad del enunciado (que se cojan las entradas) lo que está persiguiendo el interlocutor con la forma *¿no?*. Estas partículas también han sido denominadas marcadores de interacción metadiscursivos, metacomunicativos, marcadores conversacionales, evidenciadores de alteridad o marcadores de control de contacto.

Categorías gramaticales implicadas en esta operación del discurso: adverbios oracionales y formas verbales.

Algunos **ejemplos de formas lingüísticas** en las que se materializan esta operación son: *vale, de acuerdo, desde luego, por supuesto* para la intensificación; y *bueno, relativamente, posiblemente, así así, quizá, etc.* para la atenuación

A continuación presentamos tres **ejemplos en lenguaje XML**, uno de cada tipo:

1. <MD FU="Col" MET="No" DR="No" ED="No"
MOD="Truth_Booster" EVI="No" SA="No" DEX="Textual">eso s í que</MD>
2. <MD FU="No" MET="No" DR="No" ED="No"
MOD="Truth_Hedge" EVI="No" SA="No" DEX="No">probablemente</MD>
3. <MD FU="No" MET="No" DR="No" ED="No"
MOD="Interactional" EVI="No" SA="No" DEX="No">vale ?</MD>

Con el fin de reflejar la relación semántica entre las operaciones de atenuación y modalización y los otros fenómenos pragmáticos del modelo de anotación, se especificará qué es lo que se está atenuando o intensificando, si el acto de habla, si la certeza con lo enunciado, si la relación discursiva, etc.¹⁷

4.3.4. La evidencialidad

Dentro de la tradición lingüística hispánica, se llama evidencialidad al dominio semántico relacionado con la expresión de la fuente de información, y, evidencial o marcador evidencial, a la forma lingüística específica que cumple esta función. Por ejemplo, la lengua tuyuca, lengua de la familia tucanao hablada en Venezuela, presenta afijos flexivos obligatorios que se adosan a las formas verbales; en tuyuca no se puede simplemente hacer una

¹⁷DR-Hedge, DR-Booster, Interactional-Hedge, Acte-Hedge, Formulation-Hedge, Truth-Booster, Truth-Hedge, Hedge-Action, Surprise-Booster, Surprise-Hedge, Acte-Booster, Emotion-Booster, Emotion-Hedge.

afirmación sin más, sino que obligatoriamente debe elegirse un sufijo verbal que indique si el hablante sabe lo que dice porque lo ha visto, escuchado, etc. (BERMÚDEZ, 2005). La evidencialidad solamente se ha estudiado en las lenguas cuyos paradigmas verbales la marcan con un sistema cerrado de partículas. El castellano no ha gramaticalizado la evidencialidad, y solo recientemente se están haciendo estudios sobre este fenómeno en nuestra lengua (GONZÁLEZ, 2007).

Si bien es obvio que la evidencialidad es un fenómeno de carácter cognitivo, ya que es imposible conocerlo todo de primera mano, todavía no hay un acuerdo sobre qué tipos de fuentes se pueden considerar universales. Además de esto, debemos tener en cuenta que las fuentes están jerarquizadas por el nivel de autoridad que tienen en las creencias de las personas, y que esto cambia según las culturas. Así por ejemplo, en las culturas sin tradición escrita el saber transmitido oralmente tiene más prestigio que en las culturas con tradición escrita.

Actualmente, Internet evidencia un problema bastante grave relacionado con la autoridad de las fuentes de información que aparecen en la red. Para enfrentarnos a este problema debemos además diferenciar conceptualmente entre fuente de información y portal de acceso. Por ejemplo, las afirmaciones de Jose Luis Rodríguez Zapatero sobre la crisis económica pueden aparecer en el Youtube, en el Pais Digital o en una página Web personal; en cualquiera de estos tres accesos, sus enunciados respecto a la crisis económica responderían a la misma fuente de información y tendrían la misma autoridad independientemente del acceso. Actualmente, las iniciativas de Google como Books Google o GoogleAcademics, representan ejemplos de separación de información en función del tipo de texto y del género discursivo. Las tareas de Question-Answering también se verían muy beneficiadas por cualquier iniciativa de este tipo. Ya que en la red nos podemos encontrar diferentes informaciones contradictorias sobre un mismo hecho. En el futuro, la Antropología y la Sociología deben colaborar con la Lingüística para jerarquizar las fuentes de autoridad en cada cultura.

Esta es la tipología que evidenciales que hemos improvisado, para dar cuenta de los datos. Dicha tipología deberá ser sometida a revisión posterior, ya que como decimos constituye un primer acercamiento a este fenómeno en español. Hemos diferenciado las siguientes fuentes: conocimiento cultural, fuente indirecta, fuente inferencial en primera persona, fuente visual en primera persona, fuente procedente de otros sentidos, fuente cuyo origen está en otras personas, y otros.¹⁸ Esta clasificación no es muy detallada teniendo en cuenta las tipologías que se han hecho al respecto, pero diferencia dos grandes grupos, comunes a

¹⁸Cultural-Knowledge, Indirect-Discourse, Inference-FirstPerson, First-Person, Sight-FirstPerson, OtherSenses, Others-Persons.

todas. De esta manera, evitamos la polémica a nivel teórico que existe en español a la hora de clasificar más detalladamente marcadores de evidencialidad como por ejemplo *por lo visto* (GONZÁLEZ, 2007). Como nuestra clasificación es muy reduccionista, esta partícula, encaja perfectamente en la categoría de la persona que no ha presenciado este hecho, esto es, fuente indirecta. Además, este marcador evidencial tendría activado el rasgo de atenuación, ya que el hablante, al explicitar que él no es la fuente de conocimiento, elude la responsabilidad de asumir el acto de habla aseverativo. Queda así reflejada en el modelo de anotación la relación entre modalidad epistémica y evidencialidad mencionada frecuentemente por la bibliografía.

Las **categorías gramaticales** encargadas de llevar a cabo estas funciones son: sintagmas preposicionales, adverbios y sintagmas verbales.

Algunos ejemplos de **formas lingüísticas** en las que se materializan esta operación son: *por lo visto, personalmente, en mi opinión, como veis, como se suele decir, como se dice, etc.*

Observemos que algunas de estas palabras, como por ejemplo, *en mi opinión* también tendrán el rasgo Framework, que indica que el enunciado solo es verdad dentro de el contexto que marca el marcador discursivo.

A continuación presentamos **un ejemplo en lenguaje XML**:

```
<MD FU="Col" MET="No" DR="No" ED="No" MOD="No" EVI="other"
SA="No" DEX="No">supuestamente</MD>
```

4.3.5. Deixis

Por deixis ¹⁹ entendemos la referencia a elementos espacio-temporales del contexto o de la situación comunicativa en la que tiene lugar una interacción verbal.

De cara al modelo de anotación, voy a diferenciar entre dos tipos de deixis: la deixis discursiva, esto es, la referencia a otras partes del texto, y la deixis social o referencia al interlocutor.

Las **categorías gramaticales** encargadas de llevar a cabo estas funciones son: sintagmas preposicionales, adverbios y vocativos.

Algunos ejemplos de **formas lingüísticas** en las que se materializan esta operación son: *y eso, pues eso, por eso, mira, mire, señorita, señora, guapa, bonita, hija.*

A continuación presentamos unos **ejemplos en lenguaje XML**:

```
<MD FU="No" MET="No" SD="Time" TD="Enunciation"
```

¹⁹Social-Interlocutor y Textual-Contextual

DR="Coargumentation1" ED="No" MOD="No"
EVI="No" SA="No" DEX="Textual_Contextual">anteriormente</MD>

<MD FU="Col" MET="No" DR="No" ED="No"
MOD="Hedge" EVI="No" SA="No"
DEX="Social_Interlocutor">hija m ía</MD>

4.3.6. Actos de habla

La cuestión más relevante que podemos tocar en este apartado una vez ya presentadas en el capítulo anterior las bases generales de la teoría, es el tema de la tipología de los actos de habla, por un lado, y el de la relación entre fuerza ilocutiva y forma lingüística por otro; sobre todo en lo que a una perspectiva computacional de la teoría se refiere.

¿Cuántos actos de habla existen? ¿Son ilimitados? ¿Son universales? Para estas preguntas, el estado de la cuestión al respecto no expresa una postura unitaria. Ya en la época griega, Diógenes y Crisipo hacen una tipología de actos lingüísticos: preguntas, inquisiciones, órdenes, ruegos, promesas, suposiciones y problemas (SERRANO, 1983). La tipología más conocida es la de Searle, el cual distingue entre: aseverativos, compromisivos, declarativos, expresivos y directivos. Esta tipología es la que se toma de punto de referencia en las aplicaciones computacionales que intentan clasificar automáticamente actos de habla en dominios restringidos (por ejemplo, la agencia de viajes en DAMSL (ALLEN y CORE, 1997)). Otras aplicaciones hacen tipologías más *ad hoc* adecuadas a los dominios en los que van a trabajar.

Por otra parte, en cuanto a la cuestión entre forma lingüística y significado ilocutivo, los pragmáticos teóricos, a pesar de las ideas de Searle, se muestran muy tajantes respecto de la no relación entre significado ilocutivo y forma lingüística. Portolés nos explica en su libro *Pragmática para Hispanistas*, cómo un mismo enunciado *Hace calor* puede significar diferentes actos dependiendo del contexto en el que nos encontremos.

De momento, para las unidades de información pragmática que hemos tenido que marcar, **la tipología** de Searle nos ha resultado insuficiente. Nuestros tipos son: Compromiso, deseo, rechazo, aceptación, petición, aserción, asentimiento, reciprocidad, resignación, justificación, queja, resignación, apertura y cierre de saludo, gratitud, acción, advertencia y disculpa. Es obvio que no nos podemos sentir orgullosos de esta clasificación que se ha ido improvisando a medida que se iban etiquetando las diferentes unidades y que necesaria-

mente tiene que ser sistematizada en el futuro²⁰.

Ejemplos de formas lingüísticas con significado ilocutivo son: *a tomar por culo, que descanse en paz, anda ya, no te jode, no me jodas, no te digo, ánimo y vamos allá*.

A continuación presentamos un **ejemplo en lenguaje XML**:

```
<MD FU="Col" MET="No" DR="No" ED="No" MOD="No" EVI="No"  
SA="Greeting_Open" DEX="No">hola</MD>
```

4.3.7. Metáfora y Unidades Fraseológicas

Morir mártir es inyectar sangre en las venas de nuestra sociedad.

Persépolis. Marjane Satrapi.

George Lakoff y Mark Johnson pertenecen a la segunda generación de la Ciencia Cognitiva y reivindican una filosofía empíricamente responsable. Para sus investigaciones, se apoyan en los resultados de la Ciencia Cognitiva y en las teorías neuronales del lenguaje. Fundamentalmente, estos autores están interesados en la construcción del mundo por parte del hombre. Según ellos, la razón surge de la naturaleza de nuestro cerebro, del cuerpo y de la experiencia corpórea. Nuestra habilidad para hacer metáforas refleja el proceso de cognición de la realidad del ser humano: “La mayor parte de nuestro sistema conceptual se estructura metafóricamente. Conceptos tales como el amor, el tiempo, las emociones, etc. requieren una definición metafórica, ya que dichos conceptos no están claramente estructurados o delineados”(LUQUE y MANJÓN, 2002).

Una de las cuestiones que suelen tratarse en el caso de las metáforas es el tema de su naturaleza cognitivo-cultural. Para los mayas, las raíces de los árboles eran serpientes que mordían las entrañas de la tierra (MARINA, 1993). G. Lakoff, en sus estudios del inglés, el chino, el húngaro y otras lenguas, ha intentado probar que las metáforas que se asocian a la ira o al enfado están asociadas al calor y a la presión, y no al frío o a la falta de aquella. Al parecer, en todas las lenguas del mundo, existe una estructura de nuestra experiencia directamente vinculada al cuerpo humano y a la manera en que este interactúa con el entorno (LUQUE y MANJÓN, 2002). El cuerpo humano es la inmediata referencia que usan las personas para medir y medirse con el mundo (LUQUE y MANJÓN, 2002).

Ya hemos señalado los orígenes metafóricos de algunos ordenadores del discurso. Cuando esto sucede, he decidido anotarlo diferenciando tres campos: el primero explicita si la unidad de información pragmática tiene o no una base metafórica; el segundo campo indica

²⁰Commitment, Wish, Rejection, Acceptation, Ask, Assertion, Compliment, Assent, ReciprocalAction, InteractionClose, Risk, Justification, Complaint, Resignation, GreetingOpen, Refuse, GreetingClose, Gratitude, Action, Warning, Apologize

cuál es el dominio fuente utilizado, y el tercero, el dominio meta.

En Lingüística Computacional ya existen algunas iniciativas sobre el reconocimiento automático de las expresiones metafóricas (BERBER, 2007).

En nuestro corpus, sólo se van a etiquetar las metáforas de las unidades fraseológicas que aludan a estos fenómenos a los que hemos hecho referencia.

A continuación, hablaremos de la relación entre unidades fraseológicas y metáfora. Según Lakoff, los frasemas suelen ser signos motivados. Además, Kurchtkize también apunta a que los fraseologismos o modismos surgen como resultado de un proceso de metaforización (LUQUE y MANJÓN, 2002).

Desde el marco generativista, se ha insistido mucho en la perspectiva recursiva y creativa del lenguaje. Sin embargo, la recursividad no tiene por qué significar invención. Bajtin (REYES, 1990), la Lingüística de Corpus y los estudios sobre fraseología desvelan una visión de la lengua como un mecano con diferentes grados de composicionalidad. Gloria Corpas Pastor (CORPAS, 1996) señala que “la formación, el funcionamiento y el desarrollo del lenguaje están determinados no solo por reglas libres del sistema, sino también por todo tipo de estructuras prefabricadas de las que se sirven los hablantes en sus producciones lingüísticas”.

Así pues, si bien es cierto que las lenguas se componen de palabras que se combinan a través de reglas sintácticas, no es menos cierto también que hay secuencias de palabras con diferente grado de lexicalización que suelen aparecer juntas; además, no se trata solo secuencias de palabras, sino oraciones enteras que se memorizan, se almacenan y luego se sacan a colación. Así por ejemplo pensemos en qué es, desde un punto de vista cognitivo, un refrán. Un refrán es conocimiento sistematizado de la realidad que se ha codificado a través de unas imágenes pertenecientes al dominio experiencial de lo concreto y a través de una estructuras lingüísticas (entre ellas rima y el paralelismo sintáctico) que favorecerán la conservación del mismo en la memoria y la transmisión cultural. Los refranes expresan a través de imágenes concretas una generalización sobre la realidad basada en la experiencia. *Agua pasada no mueve molino* expresa con una imagen del mundo rural un conocimiento abstracto sobre la falta de causalidad de los sucesos pasados.

Para la clasificación de las unidades fraseológicas, vamos a seguir el conocido manual de Gloria Corpas Pastor (CORPAS, 1996), la cual diferencia entre colocaciones, locuciones y enunciados fraseológicos:

UFS		
-enunciado		+ enunciado
-acto de habla		+ acto de habla
fijación de la norma	fijación del sistema	fijación del habla
esfera I	esfera I	esfera III
colocaciones	locuciones	enunciados fraseológicos (refranes, dichos, sentencias)

A continuación presentamos las definiciones de cada uno de estos tres tipos:

1. **Definición de colocación:** son unidades fraseológicas que, desde el punto de vista del sistema de la lengua, son sintagmas completamente libres, generados a partir de reglas, pero que al mismo tiempo, presentan cierto grado de restricción combinatoria determinada por el uso (cierta fijación interna). Este rasgo esencial diferencia las colocaciones de las combinaciones libres de palabras.²¹
2. **Definición de locuciones:** unidades fraseológicas del sistema de la lengua con los siguientes rasgos distintivos: fijación interna, unidad de significado y fijación externa. Estas unidades no constituyen enunciados completos y generalmente funcionan como elementos oracionales.
3. **Definición de enunciados:** Las UFS de la tercera esfera son enunciados completos en sí mismos que se caracterizan por constituir actos de habla y por presentar fijación interna (material y de contenido) y externa.

La relación entre Pragmática y unidades fraseológicas en lengua española ha sido estudiada fundamentalmente por Inmaculada Penadés (PENADÉS, 1997). En su obra *Una aproximación pragmática a las unidades fraseológicas*, esta autora analiza estas lexicalizaciones a la luz del marco de la teoría de los actos de habla y de las máximas de Grice. Penadés nos muestra cómo algunas unidades fraseológicas expresan actos de habla. Así por ejemplo, las expresiones, *mal rayo te parta* o *no te conocerá ni la madre que te parió* aluden a una amenaza como acto de habla. Así como la expresión *poner la mano en el fuego* es un acto de intensificación del acto de habla de la aseveración sobre los hechos futuros *Pongo la mano en el fuego que mi mujer nunca dejará de quererme*.

Las UFS abarcan todas las *categorías gramaticales*.

Vamos a mostrar un ejemplo de una unidad fraseológica con raíz metafórica **etiquetada en XML**:

²¹La dimensión metafórica de las colocaciones fue advertida por Browning en 1967.

<MD FU="Loc" MET="S í" SD="Body" TD="Negative_Emotion" DR="No" ED="No"
MOD="No" EVI="No" SA="Rejection" DEX="No">a tomar por culo</MD>

4.4. Conclusión

En este capítulo, hemos explicado el modelo de anotación pragmática de corpus, sus presupuestos teóricos y una descripción de cada uno de sus fenómenos partiendo de su origen cognitivo-social hasta su materialización en las formas lingüísticas etiquetadas por PRAGMATEXT en lenguaje XML.

Como sabemos, una de las cualidades que confieren validez a los modelos tiene que ver con sus condiciones de replicación. Es por ello que me gustaría señalar las investigaciones realizadas con Doaa Samy ((SAMY y GONZÁLEZ-LEDESMA, 2008), (GONZÁLEZ-LEDESMA y SAMY, 2008) y (GONZÁLEZ-LEDESMA, 2007a)), en las que hemos utilizado este modelo de anotación para etiquetar el corpus paralelo de la ONU en las lenguas español, inglés y árabe. Asimismo, en el capítulo dedicado a la aplicación del modelo a la Enseñanza de español para extranjeros, se estudian también las condiciones de replicación a los marcadores discursivos del C-ORAL-ROM italiano. Por último, con este mismo modelo también hemos anotado el corpus del español oral infantil CHIEDE elaborado por la lingüista Marta Garrote Salazar (GONZÁLEZ-LEDESMA y GARROTE, 2007).

Así pues, aquí termina la explicación del modelo. Hemos considerado de rigor que el capítulo siguiente esté dedicado a la clasificación semántica de los marcadores discursivos que aparecen en nuestro corpus. Sobre el significado de estas partículas se ha escrito mucho en las últimas décadas. Existe lo que podríamos denominar una cantidad ingente de bibliografía que analiza los marcadores discursivos desde diferentes parcelas de conocimiento. La ambigüedad es uno de los retos más importantes que los marcadores del discurso plantea a La Lingüística Computacional. Tanto es así que Laura Alonso (ALONSO y otros, 2002b), a la que mencionamos en el pasado capítulo por haber realizado una base computacional de marcadores discursivos para utilizarla en resúmenes automáticos, señala que los significados de los marcadores están sometidos a sesgos teóricos. La cuestión no es baladí. Es cierto que los marcadores discursivos muestran una polifuncionalidad riquísima. También es cierto, que han sido estudiados desde diferentes disciplinas teóricas. Pero creo que se pueden encontrar puntos de vista en común, y también hacer convivir estos significados separándolos en rasgos, para que se reflejen las diferentes dimensiones. En el siguiente capítulo se abordarán estas cuestiones relativas a las carencias de sistematización descriptiva y formal de estas partículas, al tiempo que expondremos cómo hemos etiquetado cada

uno de los marcadores encontrados en el corpus.

Capítulo 5

Los marcadores discursivos de C-ORAL-ROM clasificados con PRAGMATEXT

5.1. Introducción

Antes de mostrar las tablas de clasificación de unidades de información pragmática procedentes del corpus C-ORAL-ROM, me gustaría dedicar la primera parte de este capítulo a desarrollar más detenidamente algunas cuestiones ya presentadas en la parte final del capítulo tres relacionadas con los problemas teóricos y computacionales de los marcadores discursivos.

5.2. Pragmatext y los marcadores discursivos

Ya hemos mencionado en más de una ocasión que en este trabajo hemos intentado trazar un puente de comunicación entre estas dos disciplinas, la Pragmática Lingüística y la Pragmática Computacional, en lo que al tratamiento de los marcadores discursivos se refiere. Siempre bajo la premisa de llegar a una solución de compromiso a la hora de realizar esta transferencia de conocimiento de una disciplina a la otra.

Por un lado, desde la Pragmática Lingüística, debemos considerar que los marcadores discursivos están codificando información de diferentes clases, no solo textual, también interactiva, modal e ilocutiva; por otro, también debemos tener en cuenta que algunos de ellos tienen más de un valor, y que este valor depende del género conversacional (y por tanto del tipo de corpus con el que trabajemos). A esto tenemos que añadir que algunos de ellos tienen una forma lingüística que puede cumplir otras funciones gramaticales diferentes en

el nivel de la oración, esto es, algunos de ellos tienen ambigüedad categorial. Y en cuarto y último lugar, también debemos atender al hecho de que un marcador discursivo en un mismo texto puede tener más de un valor semántico-pragmático. Ninguno de estos problemas, salvo el de la ambigüedad categorial en investigaciones muy puntuales (ZUFFEREY y POPESCU-BELIS, 2004) son reseñados por los estudios que localizan automáticamente marcadores discursivos en los textos.

Por otra parte, si nos planteamos el problema desde una perspectiva computacional, necesitamos sistematizar y formalizar los marcadores discursivos para poder operar con ellos, al tiempo que debemos considerar las limitaciones del trabajo con textos sin contexto (como son los corpus no-multimodales como este), los cuales deben ser procesados automáticamente. Si implementáramos, por ejemplo, todos los valores del marcador discursivo *en fin* que ha establecido la tradición de estudios pragmáticos tendríamos un etiquetado muy rico, pero debería ir acompañado de un etiquetado de corpus a nivel de enunciado que distinguiera la misma información, y hoy por hoy todavía no hay corpus etiquetados con este nivel de profundidad. Por tanto, debemos acercarnos a unas definiciones básicas que se mantengan constantes en los contextos de realización.

Buena parte de este compromiso al que nos referíamos antes pasa por adoptar una postura coherente con respecto a los problemas derivados del procesamiento computacional de los marcadores discursivos, y que pasamos a enumerar a continuación acompañados de la decisión que he tomado a la hora de abordarlos:

1. **Inventario de marcadores discursivos:** es momento de que la comunidad científica llegue a un acuerdo sobre qué es marcador discursivo y qué no, por un lado, y sobre el número de marcadores discursivos así como de su distribución conversacional en términos diatópicos, diastráticos y diafásicos. Los corpus deben ayudarnos a esta tarea de recopilación y documentación sobre la lengua. En este capítulo se presenta un lexicon que será accesible desde la red con posibilidad de que los usuarios lo consulten y lo amplíen libremente.
2. **Ambigüedad categorial:** muchas palabras que tienen un papel discursivo cumplen a su vez otras funciones en el nivel oracional, como nombres, adjetivos, etc. Por tanto, toda metodología de etiquetado de marcadores debe primero dividirlos en ambiguos y no ambiguos, antes de trabajar con ellos.
3. **Polifuncionalidad:** hay marcadores que cumplen diferentes funciones discursivas en función de su posición y del tipo de discurso. Por lo tanto, se deben establecer a priori

cuántos valores se van a distinguir para cada marcador, siempre teniendo en cuenta el coste de procesamiento que la diferenciación implica.

4. **Función discursiva y función oracional:** ¿debemos etiquetar solo las ocurrencias de la conjunción y cuando une oraciones y no cuando une otro tipo de cláusulas como por ejemplo sintagmas? En Pragmatext, la conjunción y siempre va a ser un conector, independientemente de las unidades que conecte, ya que nos movemos en otro nivel de procesamiento, donde un sintagma nominal también puede ser un enunciado.
5. **Idiomatidad:** Necesitamos llegar a un acuerdo sobre cómo tratar la frecuente coaparición de más de un marcador discursivo en determinados contextos, tales como *pero si*, en oraciones del tipo *Pero si yo no he sido*. Al tiempo que también debemos decidir si siguen siendo o no el mismo marcador *claro y claro que, o sea y o sea que*, etc. Nosotros hemos mantenido la sinonimia salvo en los casos en que era demasiado evidente que la construcción ya tenía entidad semántica propia. Son casos como *pues nada, y nada, y tal*, etc. En los demás casos los hemos etiquetado separados, por ejemplo, *pero si* serán dos marcadores distintos, y *claro y claro que* serán considerados el mismo marcador a la espera de estudios que defiendan una clarísima diferenciación semántica.
6. **Localización automática en el texto:** Debemos desarrollar instrumentos de desambiguación tanto categorial como discursiva que tengan una cobertura y una precisión aceptables en la anotación automática de corpus. En el capítulo siete se explicarán nuestras estrategias de etiquetado automático de marcadores del discurso en el corpus.

5.3. Etiquetado en XML

A continuación, vamos a glosar más detenidamente la formalización del modelo de anotación en formato XML; información que se adelantó en el capítulo anterior con el fin de que el lector observara la transición desde la concepción teórica hasta su anotación formal.

Sobre un marcador del discurso se explicará la siguiente información:

1. Un rasgo que señale si es una unidad fraseológica o no, y el tipo: colocación, locución o enunciado, en el caso de los refranes.
2. Nivel de ambigüedad.

3. Si posee contenido metafórico y, en el caso de que así sea, especificar los campos semánticos de origen y destino implicados.
4. Si contiene contenido emocional, y en el caso de que así sea si es negativo, positivo o de otra naturaleza (sorpresa, por ejemplo).
5. Si expresa una relación argumentativa: generalización, focalización, hipótesis, etc.
6. Si es atenuante, intensificador o interactivo.
7. Si contiene un acto de habla y especificar cuál.
8. Si posee un valor deíctico, que puede ser **textual**: anteriormente, por ejemplo, en el caso del marcador *anteriormente* en *Anteriormente os comenté que ayer no hice nada*; o puede ser **social** (*pibe, piba, tío, tía, chaval, hombre, mujer, etc.*)

A pesar de que también se ha mencionado en el capítulo anterior, me gustaría que el lector me permitiera insistir de nuevo en las bondades de un modelo basado en rasgos para el nivel pragmático. En Pragmatext, estos rasgos están separados y se pueden volver operativos o no, se pueden activar de una manera o de otra, dependiendo del marcador discursivo de que se trate. Esta es una de las ventajas del modelo de anotación, esto es, su flexibilidad a la hora de reflejar significados que remiten a fenómenos o dimensiones diferentes de la comunicación lingüística. Este sistema de rasgos semántico-pragmáticos nos permite reflejar diferente tipo de información de forma simultánea. De esta forma solucionamos a nivel teórico buena parte del problema de la ambigüedad de los marcadores discursivos, ya que cada uno de estos rasgos representa una dimensión de su posible contenido. Así por ejemplo, podemos dar cuenta de la interacción entre modalidad y evidencialidad, o de la modalidad y actos de habla, sin necesidad de crear una etiqueta o un tipo de marcador discursivo nuevo, o de crear un nuevo valor o una nueva función textual. También nos permite establecer paralelismos y contrastes sutiles entre marcadores discursivos de una misma familia. Por ejemplo, la diferencia entre *qué bien* y *menos mal*, es que ambas expresiones, que son emocionales, expresan un juicio positivo de la realidad, sin embargo, *qué bien* tiene un rasgo de intensificación que en *menos mal* no está. Otro ejemplo. El marcador discursivo *gracias a* tendrá el rasgo de relación causal activado así como el de evaluación positiva, ya que en este marcador, la causa se presenta con rasgo emocional positivo por parte del hablante.

A continuación, podemos ver un ejemplo de cómo queda formalizado este modelo de anotación en lenguaje XML.

DM	Discourse Marker
FU	Fraseological Unit
MET	Metaphore
DR	Discourse Relation
ED	Emotional Discourse
MOD	Modality
EVI	Evidentiality
SA	Speech Act
DEX	Deixis
Grupo	Nivel de ambigüedad
Tipo	Posición contextual

Tabla 5.1: Tabla de firmas de las etiquetas

<PI Grupo="1" Regla="2.8" FU="Loc" MET="No" DR="Concreteness"
ED="No" MOD="No" EVI="No"SA="No" DEX="No">por ejemplo</PI>

Esta es la información que le podemos preguntar a los textos de C-Oral-Rom sobre el marcador discursivo *por ejemplo*. Del cual se ha explicitado que es una locución (UF=No), que no contiene contenido metafórico (MET=No), que codifica una operación discursiva de concreción (DR=Concreteness), que no tiene un contenido emocional (ED=No), que no contiene modalidad (MOD=No), que no marca una fuente de conocimiento (EVI=No), que no alude a un acto de habla (SA=No), ni indica a ningún espacio del texto, ni a ninguna persona (DEX=No), que tiene un nivel de ambigüedad 1 y que su regla contextual es 2.8 ¹. En las tablas 5.1 y 5.2 se recogen todas las etiquetas de nuestro modelo de anotación.

Por razones de espacio, no se comentan aquí las razones por las que hemos decidido que cada uno de los 621 marcadores que contiene nuestro lexicón ha sido etiquetado en la manera en que aparecen en las tablas. Esta parte de la información se podrá consultar en el interfaz Web, donde aparecerá un link para cada marcador, y de cada uno se podrá obtener una justificación de por qué se ha etiquetado así en el caso de que esta decisión no sea compartida por la tradición bibliográfica; también se podrá acceder a ejemplos del corpus en los que aparezca dicho marcador, y podremos ver de nuevo su o sus etiquetas, en el caso de que tenga más de un significado. Adelantamos que hay partículas como *como* u *hombre* cuya clasificación requiere un comentario extenso, y por ello remitimos al lector al interfaz, y en cambio otras como *afortunadamente* o *como consecuencia* que son bastante intuitivas y no requieren comentario. Pasemos pues a mostrar las tablas de clasificación.

¹Estas dos últimas anotaciones las explicaremos en el capítulo siete.

FU	Col, Loc, Proverb, No
MET	Sí, No
SD	Array, Orientation, Levels-of-Space, Possibilities, Weight, Calculations, Food, Family, Human-Body, Human-Being, Mythology, Agreement, Religion, End, Existence, Sight, Action, Perception, Nature, Physiological-Activity, Movement, Hell, Milk, Levels-of-Reality, Levels-of-Deep, Body, Space, Light, Time, Attention, Calculate, Quantity, Goodness
TD	Future, Narration, The-Truth, Meaning, Ways, Opinion, Knowledge, Discourse, Argumentation, Summary, Action, Hedge, Emotion, Rejection, Intensity, Sight, Perception, Concessiveness, Demonstration, Quantity, Deduction, Agree, Surprise, Time, Apologize, Explanation, Complaint, Goodness, Negative-Evaluation, Positive-Evaluation, Levels-of-Truth, Enunciation, Evidentiality, Sincerity, Discourse-Argumentation, Interaction, Warning, Evidence, No
DR	Definite End, Cataphoric Cause, Discourse Production, Contraargumentation, Digression, Coargumentation1, Indefinite Transitional End, Coargumentation2, Coargumentation3, Reformulation, Hypothesis, Option, Conclusion Summary, Cause, Consequence, Anaphoric Concessiveness Cataphoric Conclusion, Framework, Generalization, Concreteness, Anaphoric Cause Cataphoric Consequence, Scalar Meaning, No
ED	Positive-Evaluation, Negative-Evaluation, Surprise, No
MOD	DR-Hedge, DR-Booster, Interactional-Hedge, Acte-Hedge, Formulation-Hedge, Truth-Booster, Truth-Hedge, Hedge-Action, Surprise-Booster, Surprise-Hedge, Acte-Booster, Emotion-Booster, Emotion-Hedge, Interaction
EVI	Inference, Cultural-Knowledge, Indirect-Discourse, Inference-First-Person, First-Person, Others, Sight-First-Person, Other-Senses, Others-Persons, No
SA	Commitment, Wish, Rejection, Acceptation, Ask, Assertion, Compliment, Assent, Reciprocal-Action, Interaction-Close, Risk, Justification, Complaint, Resignation, Greeting-Open, Refuse, Greeting-Close, Gratitude, Action, Warning, Apologize
DEX	Contextual, Social-Interlocutor, Textual-Contextual
Grupo	1, 2, 3
Tipo	nº de regla de desambiguación contextual

Tabla 5.2: Tabla de valores de cada atributo

5.4. Los marcadores del discurso de C-ORAL-ROM clasificados con PRAGMATEXT

5.4.1. Marcadores de Operaciones Discursivas

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
en general	Loc	No	No	No	Generalization	No	No	No	No	No
eso es como todo	Col	No	No	No	Generalization	No	No	Inference	No	No
normalmente	No	Sí	No	No	Generalization	No	No	Inference	No	No
por lo general	Loc	No	No	No	Generalization	No	No	Inference	No	No
concretamente	No	No	No	No	Concretness	No	No	No	No	No
el caso es que	Col	No	No	No	Concretness	No	No	No	Justification	No
en concreto	Loc	No	No	No	Concretness	No	No	No	No	No
en el caso de	Col	No	No	No	Concretness	No	No	No	No	No
en el caso del	Col	No	No	No	Concretness	No	No	No	No	No
en ese caso	Col	No	No	No	Concretness	No	No	No	No	No
la cosa es que	Col	No	No	No	Concretness	No	No	No	Justification	No
la cuestión es que	Col	No	No	No	Concretness	No	No	No	Justification	No
lo que pasa es que	Col	No	No	No	Concretness	No	No	No	Justification	No
lo que pasa que	Col	No	No	No	Concretness	No	No	No	Justification	No
misimamente	No	No	No	No	Concretness	No	No	No	No	No
por ejemplo	Loc	No	No	No	Concretness	No	No	No	No	No

Tabla 5.3: Relaciones Discursivas Generalization y Concretness

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
de entrada	Loc	Sí	Space	Discourse	Coargumentation1	No	No	No	No	Textual
primero	No	No	No	No	Coargumentation1	No	No	No	No	Contextual
ante todo	Loc	Sí	Space	Discourse	Coargumentation1	No	No	No	No	Contextual
antes de na-	Loc	Sí	Space	Discourse	Coargumentation1	No	No	No	No	Contextual
da	Loc	Sí	Space	Discourse	Coargumentation1	No	No	No	No	Contextual
antes que	Loc	Sí	Space	Discourse	Coargumentation1	No	No	No	No	Contextual
nada	Loc	Sí	Space	Discourse	Coargumentation1	No	No	No	No	Contextual
en primer	Loc	Sí	Array	Discourse	Coargumentation1	No	No	No	No	No
lugar	Loc	Sí	No	No	Coargumentation1	No	No	No	No	No
para em-	Loc	Sí	No	No	Coargumentation1	No	No	No	No	No
pezar	Loc	Sí	No	No	Coargumentation1	No	No	No	No	No
por una	Loc	Sí	No	No	Coargumentation1	No	No	No	No	No
parte	Loc	Sí	No	No	Coargumentation1	No	No	No	No	No
primeramente	No	No	No	No	Coargumentation1	No	No	No	No	No
aparte de	Loc	Sí	Space	Discourse	Coargumentation2	No	No	No	No	Textual
aparte	No	Sí	Space	Discourse	Coargumentation2	No	No	No	No	Contextual
de un lado	Loc	Sí	Space	Discourse	Coargumentation2	No	No	No	No	Contextual
ni	No	No	No	No	Coargumentation2	No	Truth Booster	No	No	No
por otro	Loc	Sí	Space	Discourse	Coargumentation2	No	No	No	No	Textual
por un lado	Loc	Sí	Space	Discourse	Coargumentation2	No	No	No	No	Contextual
segundo	No	No	No	No	Coargumentation2	No	No	No	No	Contextual
tercero	No	Sí	No	No	Coargumentation2	No	No	No	No	Contextual

Tabla 5.4: Relaciones Discursivas Coargumentation1, 2, 3

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
que si	Col	No	No	No	Coargumentation2	Negative Evaluation	No	No	Com-plaint	No
y	No	No	No	No	Coargumentation2	No	No	No	No	No
ni	No	No	No	No	Coargumentation2	No	Truth Booster	No	No	No
a contin- uación	Col	No	No	No	Coargumentation2	No	No	No	No	No
a su vez	Loc	No	No	No	Coargumentation2	No	No	No	No	No
además	No	No	No	No	Coargumentation2	No	No	No	No	No
aparte de	Loc	Sí	Space	Discourse	Coargumentation2	No	No	No	No	Textual
eso										Contextual
aparte de	Loc	Sí	Space	Discourse	Coargumentation2	No	No	No	No	No
que										
aparte que	Col	Sí	Space	Discourse	Coargumentation2	No	No	No	No	No
así										
las	Col	No	No	No	Coargumentation2	No	No	No	No	Textual
cosas										Contextual
asimismo	No	No	No	No	Coargumentation2	No	No	No	No	No
de la misma	Col	No	No	No	Coargumentation2	No	No	No	No	No
manera										
de manera	Loc	No	No	No	Coargumentation2	No	No	No	No	No
que										
de modo										
que	Loc	No	No	No	Coargumentation2	No	No	No	No	No
en parte	Loc	Sí	Space	TheTruth	Coargumentation2	No	TruthHedge	No	No	No
en segundo	Loc	Sí	No	No	Coargumentation2	No	No	No	No	No
lugar										
en tercer lu- gar	Loc	Sí	No	No	Coargumentation2	No	No	No	No	No
eso por una										
parte	Col	No	No	No	Coargumentation2	No	No	No	No	No
por otra										
parte	Loc	No	No	No	Coargumentation2	No	No	No	No	No

Tabla 5.4: Relaciones Discursivas Coargumentation1, 2, 3

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
por otro lado	Loc	No	No	No	Coargumentation2	No	No	No	No	No
también	No	No	No	No	Coargumentation2	No	No	No	No	No
tampoco	No	No	No	No	Coargumentation2	No	No	No	No	No
y bien	Loc	Sí	No	No	Coargumentation2	Positive Evaluation	No	No	No	No
igualmente	No	No	No	No	Coargumentation2	No	No	No	No	No
bien	No	No	No	No	Coargumentation2	Positive Evaluation	No	No	No	No
igual	No	No	No	No	Coargumentation2	No	No	No	No	No
en última instancia	Loc	Sí	No	No	Coargumentation3	No	No	No	No	No
en último término	Loc	Sí	No	No	Coargumentation3	No	No	No	No	No
por último	Loc	Sí	No	No	Coargumentation3	No	No	No	No	No
si es que es verdad	Col	No	No	No	Coargumentation3	No	Truth Booster	No	No	No
si es que eso	Col	No	No	No	Coargumentation3	No	Truth Booster	No	No	Textual Contextual

Tabla 5.4: Relaciones Discursivas Coargumentation1, 2, 3

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
ahora	No	Sí	Time	Discourse	Contraargumentation	No	No	No	No	Time
ahora bien	Col	No	No	No	Contraargumentation	No	No	No	No	No
al contrario	Col	No	No	No	Contraargumentation	No	No	No	No	No
contrariamente	No	No	No	No	Contraargumentation	No	No	No	No	No
en cambio	Loc	No	No	No	Contraargumentation	No	No	No	No	No
en caso contrario	Col	No	No	No	Contraargumentation	No	No	No	No	No
en el caso contrario de que	Col	No	No	No	Contraargumentation	No	No	No	No	No
más bien al contrario	Col	No	No	No	Contraargumentation	No	Truth Hedge	No	No	No
no obstante	Loc	No	No	No	Contraargumentation	No	No	No	No	No
pero	No	No	No	No	Contraargumentation	No	No	No	No	No
por contra	Loc	No	No	No	Contraargumentation	No	No	No	No	No
por contraste	Loc	No	No	No	Contraargumentation	No	No	No	No	No
por el contrario	Loc	No	No	No	Contraargumentation	No	No	No	No	No
sin embargo	Loc	Sí	No	No	Contraargumentation	No	No	No	No	No
sino que	Col	No	No	No	Contraargumentation	No	No	No	No	No
sólo que	Col	No	No	No	Contraargumentation	No	No	No	No	No

Tabla 5.5: Relaciones Discursivas Contraargumentation

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
con todo	Loc	Sí	Quantity	Concessiveness	Anaphoric Concessiveness Cataphoric Conclusion	No	No	No	No	No
a fin de cuentas	Loc	Sí	Calculations	Summary	Anaphoric Concessiveness Cataphoric Conclusion	No	No	No	No	No
de todas formas	Loc	Sí	Possibilities	Ways	Anaphoric Concessiveness Cataphoric Conclusion	No	No	No	No	No
de todas maneras	Loc	Sí	Possibilities	Ways	Anaphoric Concessiveness Cataphoric Conclusion	No	No	No	No	No
de todos modos	Loc	Sí	Possibilities	Ways	Anaphoric Concessiveness Cataphoric Conclusion	No	No	No	No	No
en cualquier caso	Loc	No	No	No	Anaphoric Concessiveness Cataphoric Conclusion	No	No	No	No	No
en todo caso	Loc	No	No	No	Anaphoric Concessiveness Cataphoric Conclusion	No	No	No	No	No
de ahí	Loc	Sí	Space	Discourse	Anaphoric Cause Cataphoric Conclusion sequence	No	No	No	No	Textual Contextual

Tabla 5.6: Relaciones Discursivas Anaphoric Concessiveness Cataphoric Conclusion y Anaphoric-Cause-Cataphoric-Consequence

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
de aquí	Loc	Sí	Space	Discourse	Anaphoric Cause Cataphoric Con- sequence	No	No	No	No	Textual Contextual
por ello	Col	No	No	No	Anaphoric Cause Cataphoric Con- sequence	No	No	No	No	Textual Contextual
por eso	Col	No	No	No	Anaphoric Cause Cataphoric Con- sequence	No	No	No	No	Textual Contextual
de ahí que	Col	No	No	No	Anaphoric Cause Cataphoric Con- sequence	No	No	No	No	Textual Contextual

Tabla 5.6: Relaciones Discursivas Anaphoric Concessive-
ness Cataphoric Conclusion y Anaphoric-Cause-Cataphoric-
Consequence

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
o	No	No	No	No	Option	No	No	No	No	No
o bien	Col	No	No	No	Option	No	No	No	No	No
o bien que	Col	No	No	No	Option	No	No	No	No	No
o no ?	Col	No	No	No	Option	No	TruthHedge	No	No	SocialInterlocutor
no ?	Col	No	No	No	Option	No	TruthHedge	No	No	SocialInterlocutor
o qué ?	Col	No	No	No	Option	No	TruthHedge	No	No	SocialInterlocutor
o sí ?	Col	No	No	No	Option	No	TruthHedge	No	No	SocialInterlocutor

Tabla 5.7: Relaciones Discursivas Option

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
a lo mejor	Col	No	No	No	Hypothesis	Positive Evaluation	DRHedge	No	No	No
a lo peor	Col	No	No	No	Hypothesis	Negative Evaluation	DRHedge	No	No	No
a menos que	Loc	No	No	No	Hypothesis	No	DRHedge	No	No	No
a no ser que	Loc	No	No	No	Hypothesis	No	DRHedge	No	No	No
a ver si	Col	No	No	No	Hypothesis	No	No	No	No	No
acaso	No	No	No	No	Hypothesis	No	No	No	No	No
como si	Col	No	No	No	Hypothesis	No	DRHedge	No	Assertion	No
con tal de	Loc	No	No	No	Hypothesis	No	No	No	No	No
con tal de que	Loc	No	No	No	Hypothesis	No	No	No	No	No
mira que si	Loc	Sí	No	No	Hypothesis	No	No	No	No	No
por si acaso	Loc	No	No	No	Hypothesis	No	No	No	No	No
si	No	No	No	No	Hypothesis	No	TruthHedge	No	Assertion	No
si acaso	Col	No	No	No	Hypothesis	No	TruthHedge	No	Assertion	No
siempre y cuando	Col	No	No	No	Hypothesis	No	No	No	No	No
como	No	No	No	No	Hypothesis	No	No	No	No	No
igual	No	No	No	No	Hypothesis	No	No	No	No	No

Tabla 5.8: Relaciones Discursivas Hypothesis

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
gracias a	Col	No	No	No	Cause	Positive Evalua- tion	No	No	No	No
ya que	Loc	No	No	No	Cause	No	No	No	No	No
dado que	Col	No	No	No	Cause	No	No	No	No	No
debido a que	Col	No	No	No	Cause	No	No	No	No	No
por eso mismo	Loc	No	No	No	Cause	No	Truth Booster Truth Booster	No	No	No
por eso te digo	Col	No	No	No	Cause	No	No	No	No	No
por este motivo	Col	No	No	No	Cause	No	No	No	No	No
porque	No	No	No	No	Cause	No	No	No	No	No
puesto que	Loc	No	No	No	Cause	No	No	No	No	No
y no por ello	Col	No	No	No	Cause	No	No	No	No	No
como	No	No	No	No	Cause	No	No	No	No	No
resulta que	Loc	No	No	No	CataphoricCause	No	No	No	No	No
es que	Loc	Sí	Exis- tence	Ex- pla- na- tion	Cause	No	No	No	Justi- fica- tion	No
lo que pasa	Col	No	No	No	Cause	Justi- fica- tion	No	No	No	No

Tabla 5.9: Relaciones Discursivas Cause y Cathaphoric-Cause

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
así que	Col	No	No	No	Consequence	No	No	No	No	No
así que ...	Col	No	No	No	Consequence	No	No	No	No	No
así que bueno	Col	No	No	No	Consequence	No	DRHedge	No	No	No
así que eso	Col	No	No	No	Consequence	No	No	No	No	Textual
así que mira	Col	No	No	No	Consequence	No	No	No	No	Contextual
así que nada	Col	No	No	No	Consequence	No	DRHedge	No	No	Social
así que no sé	Col	No	No	No	Consequence	No	DRHedge	No	No	Interlocutor
así que por eso te digo	Col	No	No	No	Consequence	No	DR Booster	No	No	No
así que ya te digo	Col	No	No	No	Consequence	No	DR Booster	No	No	Textual
con lo cual	Loc	No	No	No	Consequence	No	No	No	No	Contextual
en consecuencia	Loc	No	No	No	Consequence	No	No	No	No	No
entonces	No	No	No	No	Consequence	No	No	No	No	No
entonces ?	No	No	No	No	Consequence	No	No	No	No	No
por lo cual	Loc	No	No	No	Consequence	No	Interaction	No	No	Social
por lo tanto	Loc	No	No	No	Consequence	No	No	No	No	Interlocutor
por tanto	Loc	No	No	No	Consequence	No	No	No	No	No
pues	No	No	No	No	Consequence	No	No	No	No	No
pues es lo que te digo	Col	No	No	No	Consequence	No	Truth Booster	No	No	No
luego	No	No	No	No	Consequence	No	No	No	No	No

Tabla 5.10: Relaciones Discursivas Consequence

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
conclusión	No	No	No	No	Conclusion Summary	No	No	No	No	No
en fin	Loc	Sí	End	Discourse	Conclusion Summary	No	No	No	Resignation	No
en total	Loc	Sí	Calculate	Discourse	Conclusion Summary	No	No	No	No	No
muy breve	Col	No	No	No	Conclusion Summary	No	No	No	No	No
a tomar por culo	Loc	Sí	Body	Rejection	Conclusion Summary	Negative Evaluation	No	No	No	Con- tex- tual
al fin y al cabo	Loc	Sí	Space	Argumentation	Conclusion Summary	No	No	No	No	No
al final	Loc	Sí	Space	Argumentation	Conclusion Summary	Positive Evaluation	No	No	No	No
en conclusión	Loc	No	No	No	Conclusion Summary	No	No	No	No	No
en una palabra	Loc	Sí	No	No	Conclusion Summary	No	No	No	No	No
fundamentalmente	No	No	No	No	Conclusion Summary	No	No	No	No	No
grosso modo	Loc	No	No	No	Conclusion Summary	No	No	No	No	No
principalmente	No	No	No	No	Conclusion Summary	No	No	No	No	No
resumiendo un poco	Col	No	No	No	Conclusion Summary	No	No	No	No	No
vamos a decirlo muy brevemente	Col	No	No	No	Conclusion Summary	No	No	No	No	No

Tabla 5.11: Relaciones Discursivas ConclusionSummary

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
digamos	No	No	No	No	Reformulation	No	TruthHedge	No	No	No
esto es	Col	No	No	No	Reformulation	No	No	No	No	Textual
quiero decir	Col	No	No	No	Reformulation	No	TruthHedge	No	No	Contextual
vamos	No	No	No	No	Reformulation	No	TruthHedge	No	No	No
a saber	Loc	No	No	No	Reformulation	No	No	No	No	No
a ver si me expli-	Col	No	No	No	Reformulation	No	No	No	No	No
co										
dicho de otro mo-	Col	No	No	No	Reformulation	No	No	No	No	No
do										
es decir	Loc	No	No	No	Reformulation	No	No	No	No	No
mejor dicho	Loc	No	No	No	Reformulation	No	No	No	No	No
miento	No	No	No	No	Reformulation	No	TruthHedge	No	No	No
o lo que es lo mis-	Col	No	No	No	Reformulation	No	No	No	No	No
mo										
o sea	Loc	No	No	No	Reformulation	No	No	No	No	No
por así decirlo	Col	No	No	No	Reformulation	No	TruthHedge	No	No	No
qué digo	Col	No	No	No	Reformulation	No	TruthHedge	No	No	No
quiero decir que	Col	No	No	No	Reformulation	No	TruthHedge	No	No	No
si me explico	Col	No	No	No	Reformulation	No	No	No	No	No
digo	No	No	No	No	Reformulation	No	No	No	No	No

Tabla 5.12: Relaciones Discursivas Reformulation

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
y se ha acabado	Col	Sí	No	No	Definite End	No	No	No	No	No
y punto	Col	No	No	No	Definite End	Negative Evaluation	No	No	No	No
nada	No	No	No	No	Indefinite Transition- al End	No	Formulation Hedge	No	No	No
no sé cuántos	Col	No	No	No	Indefinite Transition- al End	No	Formulation Hedge	No	No	No
no sé qué no sé cuántos	Col	No	No	No	Indefinite Transition- al End	No	Formulation Hedge	No	No	No
tal	No	No	No	No	Indefinite Transition- al End	No	No	No	No	No
total	No	No	No	No	Indefinite Transition- al End	No	No	No	No	No
y demás	Col	No	No	No	Indefinite Transition- al End	No	No	No	No	Textual Contextual
y eso	Col	No	No	No	Indefinite Transition- al End	No	Formulation Hedge	No	No	Textual Contextual
y nada	Col	No	No	No	Indefinite Transition- al End	No	Formulation Hedge	No	No	Textual Contextual
y todo	Col	No	No	No	Indefinite Transition- al End	No	Formulation Hedge	No	No	Textual Contextual

Tabla 5.13: Relaciones Discursivas: Defined End e Indefinite Transitional End

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
y ya	Col	No	No	No	Indefinite Transition- al End	No	Formulation Hedge	No	No	No
etcétera	No	No	No	No	Indefinite Transition- al End	No	No	No	No	No
no sé qué historia	Col	No	No	No	Indefinite Transition- al End	No	No	No	No	No
no sé qué rollo	Col	Sí	No	No	Indefinite Transition- al End	No	No	No	No	No
o algo	Col	No	No	No	Indefinite Transition- al End	No	Truth Hedge	No	No	No
o algo de eso	Col	No	No	No	Indefinite Transition- al End	No	Truth Hedge	No	No	No
o algo por el esti- lo	Col	No	No	No	Indefinite Transition- al End	No	TruthHedge	No	No	No
pues nada	Col	No	No	No	Indefinite Transition- al End	No	No	No	No	No
que tal y cual	Col	No	No	No	Indefinite Transition- al End	No	No	No	No	No
que tal y que cual	Col	No	No	No	Indefinite Transition- al End	No	No	No	No	No

Tabla 5.13: Relaciones Discursivas: Defined End e Indefinite Transitional End

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
tal y cual	Col	No	No	No	Indefinite Transition- al End	No	No	No	No	No
toda la leche	Loc	Sí	No	No	Indefinite Transition- al End	No	No	No	No	No
total que	Loc	Sí	Cal- cula- tions	Na- rra- tion	Indefinite Transition- al End	No	No	No	No	No
y tal	Col	No	No	No	Indefinite Transition- al End	No	No	No	No	No
y tal y cual	Col	No	No	No	Indefinite Transition- al End	No	No	No	No	No
y toda la leche	Col	Sí	No	No	Indefinite Transition- al End	No	No	No	No	No
y todo ese lío	Col	Sí	No	No	Indefinite Transition- al End	No	No	No	No	No
y todo eso	Col	No	No	No	Indefinite Transition- al End	No	No	No	No	No
y ya está	Col	No	No	No	Indefinite Transition- al End	No	No	No	No	No

Tabla 5.13: Relaciones Discursivas: Defined End e Indefi-
nite Transitional End

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
de acuerdo con	Col	No	No	No	Framework	No	Truth-Booster	No	No	No
desde el punto de vista	Col	Sí	Sight	Opinion	Framework	No	No	No	No	No
desde el punto de vista de	Col	Sí	Sight	Opinion	Framework	No	No	No	No	No
desde mi punto de vista	Col	Sí	Sight	Opinion	Framework	No	Truth-Hedge	First-Person	No	No
en tanto que	Loc	No	No	No	Framework	No	No	No	No	No
desde nuestro punto de vista	Col	Sí	Sight	Opinion	Framework	No	Truth-Hedge	First-Person	No	No
en el sentido de que	Col	Sí	Orientation	Meaning	Framework	No	No	No	No	No
en este sentido	Col	Sí	Orientation	Meaning	Framework	No	No	No	No	No
en mi opinión	Col	No	No	No	Framework	No	Truth-Hedge	First-Person	No	No
vamos a ver	Loc	Sí	Sight	De-mons-tration	Discourse-Production	No	No	No	No	Social-Inter-locutor
eh	No	No	No	No	Discourse-Production	No	No	No	No	No
a propósito	Loc	No	No	No	Digression	No	No	No	No	No
a propósito de	Loc	No	No	No	Digression	No	No	No	No	No
a todo esto	Loc	No	No	No	Digression	No	No	No	No	No
por cierto	Loc	No	No	No	Digression	No	No	No	No	No
por no hablar	Col	No	No	No	Digression	No	No	No	No	No
encima	No	Sí	Space	Discourse	ScalarMeaning	Negative-Evaluation	No	No	No	No

Tabla 5.14: Relaciones Discursivas: Framework, Discourse Production, Digression, Scalar Meaning

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
casi	No	No	No	No	Scalar Meaning	No	No	No	No	No
inclusive	No	No	No	No	Scalar Meaning	No	No	No	No	No
incluso	No	No	No	No	Scalar Meaning	No	No	No	No	No
más aún	Col	No	No	No	Scalar Meaning	No	No	No	No	No
ni siquiera	Col	No	No	No	Scalar Meaning	No	No	No	No	No

Tabla 5.14: Relaciones Discursivas: Framework, Discourse
Production, Digression, Scalar Meaning

5.4.2. Marcadores de Evaluación Emocional

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
es espectacular	Col	No	No	No	No	Positive Evaluation	No	No	No	No
espectacular	No	No	No	No	No	Positive Evaluation	No	No	No	No
estupendo	No	No	No	No	No	Positive Evaluation	No	No	No	No
genial	No	No	No	No	No	Positive Evaluation	No	No	No	No
gracias a	Col	No	No	No	Cause	Positive Evaluation	No	No	No	No
guay	No	No	No	No	No	Positive Evaluation	No	No	No	No
la hostia	Loc	Sí	Religion	Intensity	No	Positive Evaluation	Emotion Booster	No	No	No
maravilloso	No	No	No	No	No	Positive Evaluation	No	No	No	No
muy bien	Col	No	No	No	No	Positive Evaluation	No	No	No	No
qué bien	Col	No	No	No	No	Positive Evaluation	No	No	No	No
superbien	No	No	No	No	No	Positive Evaluation	No	No	No	No
está bien	No	No	No	No	No	Positive Evaluation	No	No	No	No
¡hum!	No	No	No	No	No	Positive Evaluation	No	No	No	No
¡ja!	No	No	No	No	No	Positive Evaluation	No	No	No	No
¡ji!	No	No	No	No	No	Positive Evaluation	No	No	No	No
¡jo!	No	No	No	No	No	Positive Evaluation	No	No	No	No

Tabla 5.15: Marcadores de significado emocional, Positive Evaluation

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
¡ole!	No	No	No	No	No	Positive Evaluation	Emotion Booster	No	Com- pli- ment	No
¡yeah!	No	No	No	No	No	Positive Evaluation	No	No	No	No
a dios gracias	Col	No	Nature	No	No	Positive Evaluation	No	No	No	No
a lo mejor	Col	No	No	No	Hy- po- thesis	Positive Evaluation	DR Hedge	No	No	No
al final	Loc	Sí	Space	Ar- gu- men- tation	Con- clu- sion- Su- mma- ry	Positive Evaluation	No	No	No	No
alucinante	No	No	No	No	No	Positive Evaluation	No	No	No	No
bien es cierto	Col	No	No	No	No	Positive Evaluation	Emotion Booster	No	No	No
bien es cierto que	Col	No	No	No	No	Positive Evaluation	Emotion Booster	No	No	No
es chévere	Col	No	No	No	No	Positive Evaluation	No	No	No	No
es maravilloso	Col	No	No	No	No	Positive Evaluation	No	No	No	No
eso es de puta madre	Col	No	No	No	No	Positive Evaluation	No	No	No	No
eso es alucinante	Col	No	No	No	No	Positive Evaluation	No	No	No	No
eso es chévere	Col	No	No	No	No	Positive Evaluation	No	No	No	No

Tabla 5.15: Marcadores de significado emocional, Positive Evaluation

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
eso es la leche	Col	No	No	No	No	Positive	No	No	No	No
eso está bien	Col	No	No	No	No	Evaluation Positive	No	No	No	No
eso está guay	Col	No	No	No	No	Evaluation Positive	No	No	No	No
está de puta madre	Col	No	No	No	No	Evaluation Positive	No	No	No	No
está guay	Col	No	No	No	No	Evaluation Positive	No	No	No	No
está muy guay	Col	No	No	No	No	Evaluation Positive	No	No	No	No
está súper bien	Col	No	No	No	No	Evaluation Positive	No	No	No	No
gracias a Dios	Loc	No	No	No	No	Evaluation Positive	No	No	No	No
ja	No	No	No	No	No	Evaluation Positive	No	No	No	No
la mar de bien	Loc	No	No	No	No	Evaluation Positive	Emotion Booster	No	No	No
la verdad es que está muy guay	Col	No	No	No	No	Evaluation Positive	TruthBooster	No	No	No
más bien que	Loc	No	No	No	No	Evaluation Positive	TruthHedge	No	No	No
me parece estupendo	Col	No	No	No	No	Evaluation Positive	No	No	No	No
menos mal	Loc	No	No	No	No	Evaluation Positive	No	No	No	No
menos mal que	Loc	No	No	No	No	Evaluation Positive	No	No	No	No
por fin	Loc	No	No	No	Time	Evaluation Positive	No	No	No	No
por fortuna	Loc	No	No	No	No	Evaluation Positive	No	No	No	No

Tabla 5.15: Marcadores de significado emocional, Positive Evaluation

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
por suerte	Loc	No	No	No	No	Positive Evaluation	No	No	No	No
pues qué bien	Col	No	No	No	No	Positive Evaluation	No	No	No	No
que alucinas	Col	Sí	No	No	No	Positive Evaluation	No	No	No	No
qué gracia	Col	Sí	No	No	No	Positive Evaluation	No	No	No	No
qué guay	Col	No	No	No	No	Positive Evaluation	No	No	No	No
todo guay	Col	No	No	No	No	Positive Evaluation	No	No	No	No
y bien	Loc	Sí	No	No	Co-argu-men-ta-tion2	Positive Evaluation	No	No	No	No
¡joder!	No	No	No	No	No	Positive Evaluation	No	No	No	No
bien	No	No	No	No	Co-argu-men-ta-tion2	Positive Evaluation	No	No	No	No
bien	No	No	No	No	No	Positive Evaluation	No	No	No	No
joder	No	No	No	No	No	Positive Evaluation	No	No	No	No

Tabla 5.15: Marcadores de significado emocional, Positive Evaluation

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
encima	No	Sí	Space	Discourse	Scalar Meaning	Negative Evaluation	No	No	No	No
no te digo	Loc	Sí	Action	Rejection	No	Negative Evaluation	Truth Booster	No	Refuse	No
por Dios	Loc	Sí	Religion	Emotion	No	Negative Evaluation	No	No	No	No
una mierda	No	No	No	No	No	Negative Evaluation	No	No	Refuse	No
vaya	No	No	No	No	No	Negative Evaluation	No	No	No	No
que si	Col	No	No	No	Co-argu-ment-tion2	Negative Evaluation	No	No	Com-plaint	No
¡ay madre!	No	Sí	Family	Surprise	No	Negative Evaluation	No	No	No	No
¡ay!	No	No	No	No	No	Negative Evaluation	No	No	No	No
¡bah!	No	No	No	No	No	Negative Evaluation	Truth Hedge	No	No	No
¡buah!	No	No	No	No	No	Negative Evaluation	Truth Hedge	No	No	No
¡cachis en la mar!	No	Sí	Nature	Complaint	No	Negative Evaluation	No	No	No	No
¡Dios mío de mi alma!	Col	No	Nature	No	No	Negative Evaluation	No	No	No	No
¡hala!	No	No	No	No	No	Negative Evaluation	No	No	Action	No
¡huy!	No	No	No	No	No	Negative Evaluation	No	No	No	No

Tabla 5.16: Marcadores de significado emocional Negative Evaluation

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
¡jobar!	No	No	No	No	No	Negative Evaluation	Emotion Hedge	No	No	No
¡jolín!	No	No	No	No	No	Negative Evaluation	Emotion Hedge	No	No	No
¡jolines!	No	No	No	No	No	Negative Evaluation	Emotion Hedge	No	No	No
¡jope!	No	No	No	No	No	Negative Evaluation	Emotion Hedge	No	No	No
¡leche!	No	Sí	Food	Complaint	No	Negative Evaluation	Emotion Hedge	No	No	No
¡por Dios!	Loc	Sí	Religion	Emotion	No	Negative Evaluation	No	No	No	No
¡uf!	No	No	No	No	No	Negative Evaluation	No	No	No	No
¡ups!	No	No	No	No	No	Negative Evaluation	No	No	No	No
¡uy!	No	No	No	No	No	Negative Evaluation	No	No	No	No
¡va!	No	No	Nature	No	No	Negative Evaluation	No	No	No	No
¡vamos!	No	Sí	Movement	Action	No	Negative Evaluation	Truth Booster	No	No	No
a lo peor	Col	No	No	No	Hypothesis	Negative Evaluation	DR Hedge	No	No	No
a pesar	Loc	Sí	Weight	Argumentation	Concessiveness	Negative Evaluation	No	No	No	No
a pesar de	Loc	Sí	Weight	Argumentation	Concessiveness	Negative Evaluation	No	No	No	No

Tabla 5.16: Marcadores de significado emocional Negative Evaluation

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
a pesar de que	Loc	Sí	Weight	Argumentation Rejection	Concessiveness	Negative Evaluation	No	No	No	No
a tomar por culo	Loc	Sí	Body		Conclusion Summary	Negative Evaluation	No	No	No	Contextual
bah	No	No	No	No	No	Negative Evaluation	Truth Hedge	No	No	No
coño	No	Sí	Body	Emotion	No	Negative Evaluation	No	No	No	No
cuando	No	No	No	No	Time	Negative Evaluation	No	No	No	No
es lo peor	Col	No	No	No	No	Negative Evaluation	No	No	No	No
hala	No	No	No	No	No	Negative Evaluation	No	No	Action	No
hasta los cojones	Loc	No	No	No	No	Negative Evaluation	No	No	No	No
jolín	No	No	No	No	No	Negative Evaluation	Emotion Hedge	No	No	No
la verdad es que no mola mucho	Col	No	No	No	No	Negative Evaluation	Truth	No	No	No
me cago en diona	Loc	Sí	Religion	Emotion	No	Negative Evaluation	Booster	No	No	No
me cago en Dios	Loc	Sí	Religion	Emotion	No	Negative Evaluation	No	No	No	No
no me jodas	Loc	Sí	No	No	No	Negative Evaluation	No	No	Rejection	No
no te jode	Loc	Sí	No	No	No	Negative Evaluation	ActeHedge	No	Rejection	No

Tabla 5.16: Marcadores de significado emocional Negative Evaluation

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
pese a que	Loc	No	No	No	Con- cessi- ve- ness	Negative Evaluation	No	No	No	No
por desgracia	Loc	No	No	No	No	Negative Evaluation	No	No	No	No
qué coño mal	Col	Sí	No	No	No	Negative Evaluation	Truth Hedge	No	No	No
que le den por cu- lo	Loc	Sí	Body	Re- jec- tion	No	Negative Evaluation	No	No	Rejection	No
que no veas	Loc	Sí	Sight	Quantity	Quantity	Negative Evaluation	DR Booster	No	No	No
qué te vas a acor- dar	Col	No	No	No	No	Negative Evaluation	Truth Hedge	No	No	So- cial Inter- locu- tor
y punto	Col	No	No	No	Definite End	Negative Evaluation	No	No	No	No
¡coño!	No	Sí	Human Body	Emotion	No	Negative Evaluation	Emotion Booster	No	No	No
coño	No	Sí	Human Body	Emotion	No	Negative Evaluation	Emotion Booster	No	No	No
¡joder!	No	No	No	No	No	Negative Evaluation	No	No	No	No
joder	No	No	No	No	No	Negative Evaluation	No	No	No	No

Tabla 5.16: Marcadores de significado emocional Negative Evaluation

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
es muy fuerte	Col	Sí	No	No	No	Surprise	No	No	No	No
fuerte	No	No	No	No	No	Surprise	No	No	No	No
hostia	No	Sí	Religion	Surprise	No	Surprise	No	No	No	No
hostias	No	Sí	Religion	Surprise	No	Surprise	No	No	No	No
muy fuerte	Col	No	No	No	No	Surprise	No	No	No	No
no me digas	Col	No	No	No	No	Surprise	No	No	No	No
¡ah!	No	No	No	No	No	Surprise	No	No	No	No
¡ahí va!	Loc	Sí	Action	Surprise	No	Surprise	No	No	No	No
¡hostia!	No	No	No	No	No	Surprise	No	No	No	No
¡hostias!	No	No	No	No	No	Surprise	No	No	No	No
¡madre del amor hermoso!	Loc	Sí	Family	Surprise	No	Surprise	Surprise Booster	No	No	No
¡madre mía!	Loc	Sí	Family	Surprise	No	Surprise	Surprise Booster	No	No	No
¡oh!	No	No	No	No	No	Surprise	No	No	No	No
¡ostras!	No	No	Nature	No	No	Surprise	No	No	No	No
ah	No	No	No	No	No	Surprise	No	No	No	No
de la hostia	Loc	Sí	Religion	Intensity	No	Surprise	Surprise Booster	No	No	No
eso es muy fuerte	Col	No	No	No	No	Surprise	No	No	No	No
esto ya es muy fuerte	Col	Sí	No	No	No	Surprise	Emotion Booster	No	No	No
madre mía	Loc	Sí	No	No	No	Surprise	No	No	No	No
madre mía de mi vida	Loc	Sí	No	No	No	Surprise	Surprise Booster	No	No	No
no me lo puedo creer	Col	No	No	No	No	Surprise	Surprise Booster	No	No	No
oh	Col	No	No	No	No	Surprise	No	No	No	No
ostras	No	No	No	No	No	Surprise	No	No	No	No

Tabla 5.17: Marcadores de significado emocional Surprise

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
qué barbaridad	Col	Sí	No	No	No	Surprise	Surprise Booster	No	No	No
qué fuerte	Col	No	No	No	No	Surprise	No	No	No	No
¡coño!	No	Sí	Human Body	Emotion	No	Surprise	Emotion Booster	No	No	No
coño	No	Sí	Human Body	Emotion	No	Surprise	Emotion Booster	No	No	No
¡hombre!	No	Sí	Human Be- ing	Surprise	No	Surprise	No	No	No	So- cial Inter- locu- tor
hombre	No	Sí	Human Be- ing	Surprise	No	Surprise	No	No	No	So- cial Inter- locu- tor
joder	No	No	No	No	No	Surprise	No	No	No	No

Tabla 5.17: Marcadores de significado emocional Surprise

5.4.3. Marcadores de Modalización discursiva

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
así es	Col	No	No	No	No	No	Truth	No	No	Textual-Contextual
claro	No	Sí	Light	Evidence	No	No	Booster Truth	No	No	No
de acuerdo con	Col	No	No	No	Framework	No	Booster Truth	No	No	No
de acuerdo	Loc	No	No	No	No	No	Booster Truth	No	No	No
de cierto	Col	No	No	No	No	No	Booster Truth	No	No	No
de corazón	Col	Sí	Body	Sincerity	No	No	Booster Truth	No	No	No
de verdad	Col	No	No	No	No	No	Booster Truth	No	No	No
en verdad	Loc	No	No	No	No	No	Booster Truth	No	No	No
eso	No	No	No	No	No	No	Booster Truth	No	No	Social Inter-locutor
evidente	No	No	No	No	No	No	Booster Truth	No	No	No
exactamente	No	No	No	No	No	No	Booster Truth	No	No	No
la verdad	Col	No	No	No	No	No	Booster Truth	No	No	No
lo que te digo	Col	No	No	No	No	No	Truth	First Person	No	No
ni	No	No	No	No	Coargumentation2	No	Booster Truth	No	No	No
no te digo	Loc	Sí	Action	Rejection	No	Negative-Evaluation	Booster Truth	No	Refuse	No
pues eso	Loc	No	No	No	No	No	Booster Truth	No	No	Textual-Contextual
seguro	No	No	No	No	No	No	Booster Truth	No	No	No

Tabla 5.18: Marcadores de modalización: Truth booster, DR

Booster, Emotion Booster, Acte Booster y Surprise Booster

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
vale	No	No	No	No	No	No	Truth	No	No	No
para nada	Loc	No	No	No	No	No	Booster Truth	No	No	No
particular- mente	No	No	No	No	No	No	Booster Truth	No	No	No
ni	No	No	No	No	Coargu- mentation2	No	Booster Truth	No	No	No
¡vamos!	No	Sí	Movement	Action	No	Negative- Evaluation	Booster Truth	No	No	No
a decir ver- dad absolu- tamente	Col	No	No	No	No	No	Booster Truth	No	No	No
casi seguro	Col	mNo	No	No	No	No	Booster Truth	No	No	No
ciertamente	No	No	No	No	No	No	Booster Truth	No	No	No
claramente	No	No	Nature	No	No	No	Booster Truth	No	No	No
claro que no	Col	Sí	Light	Knowledge	No	No	Booster Truth	No	No	No
claro que sí	Col	Sí	Light	Knowledge	No	No	Booster Truth	No	No	No
como es natural	Col	No	No	No	No	No	Booster Truth	Cultural- Knowledge	No	No
con toda se- guridad	Col	No	No	No	No	No	Truth	No	No	No
de hecho	Loc	No	No	No	No	No	Booster Truth	No	No	No
desde luego	Loc	No	No	No	No	No	Booster Truth	No	No	No
efectiva- mente	No	No	No	No	No	No	Booster Truth	No	No	No

Tabla 5.18: Marcadores de modalización: Truth booster, DR
Booster, Emotion Booster, Acte Booster y Surprise Booster

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
en absoluto	Loc	No	No	No	No	No	Truth	No	No	No
en efecto	Loc	No	No	No	No	No	Booster Truth	No	No	No
en el fondo	Loc	Sí	Levelsof- Space	Levelsof- Truth	No	No	Truth	No	No	No
en realidad	Loc	No	No	No	No	No	Booster Truth	No	No	No
es cierto	Col	No	No	No	No	No	Booster Truth	No	No	No
es cierto que	Col	No	No	No	No	No	Truth	No	No	No
es evidente que	Col	No	No	No	No	No	Booster Truth	No	No	No
es muy probable	Col	No	No	No	No	No	Truth	No	No	No
es obvio	Col	No	No	No	No	No	Booster Truth	No	No	No
es verdad	Col	No	No	No	No	No	Booster Truth	No	No	No
es verdad que	Col	No	No	No	No	No	Truth	No	No	No
eso es cier- to	Col	No	No	No	No	No	Booster Truth	No	No	No
eso es evi- dente	Col	No	No	No	No	No	Truth	No	No	No
eso es ob- vio	Col	No	No	No	No	No	Booster Truth	No	No	No
está claro eso	Col	No	No	No	No	No	Booster Truth	No	No	No
está muy claro	Col	No	No	No	No	No	Truth Booster	No	No	No

Tabla 5.18: Marcadores de modalización: Truth booster, DR
Booster, Emotion Booster, Acte Booster y Surprise Booster

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
eso mismo	Col	No	No	No	No	No	Truth	No	No	SocialInter-locutor
eso no es cierto	Col	No	No	No	No	No	Booster Truth	No	No	No
eso sí que	Col	No	No	No	No	No	Booster Truth	No	No	No
eso también	Col	No	No	No	No	No	Booster	No	No	No
bién es	Col	No	No	No	No	No	Truth Booster	No	No	No
verdad es										
eso también										
bién está su- perclaro	Col	No	No	No	No	No	Truth Booster	No	No	No
especial- mente	No	No	No	No	No	No	Truth Booster	No	No	No
está claro	Col	No	No	No	No	No	Truth Booster	No	No	No
que										
evidente- mente	No	No	No	No	No	No	Truth Booster	No	No	No
indefectible- mente	No	No	No	No	No	No	Truth Booster	No	No	No
indudable- mente	No	No	No	No	No	No	Truth Booster	No	No	No
inevitable- mente	No	No	No	No	No	No	Truth Booster	No	No	No
justamente	No	No	No	No	No	No	Truth Booster	No	No	No
la verdad es	Col	No	No	No	No	No	Truth Booster	No	No	No
que										
la verdad es	Col	No	No	No	No	No	Truth Booster	No	No	No
que es eso										

Tabla 5.18: Marcadores de modalización: Truth booster, DR
Booster, Emotion Booster, Acte Booster y Surprise Booster

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
la verdad es que está muy guay	Col	No	No	No	No	Positive-Evaluation	Truth Booster	No	No	No
la verdad es que no	Col	No	No	No	No	No	Truth Booster	No	No	No
la verdad es que no mucho	Col	No	No	No	No	Negative-Evaluation	Truth Booster	No	No	No
la verdad es que sí	Col	No	No	No	No	No	Truth Booster	No	No	No
lo cierto es que	Col	No	No	No	No	No	Truth Booster	No	No	No
lo que está claro es que	Col	Sí	Sight	The Truth	No	No	Truth Booster	No	No	No
lo que sí es cierto	Col	No	No	No	No	No	Truth Booster	No	No	No
lo que te digo	Col	No	No	No	No	No	Truth Booster	FirstPerson	No	No
lógicamente	No	No	No	No	No	No	Truth	Inference	No	No
muy claramente	Col	No	No	No	No	No	Booster Truth	No	No	No
mente	Col	No	No	No	No	No	Booster Truth	No	No	No
nada en absoluto	Col	No	No	No	No	No	Booster Truth	No	No	No
naturalmente	No	No	No	No	No	No	Booster Truth	No	No	No
necesariamente	No	No	No	No	No	No	Booster Truth	No	No	No
ni nada	Col	No	No	No	No	No	Booster Truth	No	No	No

Tabla 5.18: Marcadores de modalización: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
ni nada de eso	Col	No	No	No	No	No	Truth-	No	No	No
obviamente	No	No	No	No	No	No	Booster Truth	No	No	No
obviamente	Col	No	No	No	No	No	Booster Truth	No	No	No
no por eso mismo	Loc	No	No	No	Cause	No	Booster Truth	No	No	No
por eso te digo	Col	No	No	No	Cause	No	Booster Truth	No	No	No
por fuerza	Loc	No	No	No	No	No	Booster Truth	Inference	No	No
por huevos	Loc	No	No	No	No	No	Booster Truth	Inference	No	No
por los cuatro costados	Loc	No	No	No	No	No	Truth Booster	No	No	No
por supuesto	Loc	No	No	No	No	No	Truth Booster	No	No	No
supuesto que sí	Col	Sí	No	No	No	No	Truth- Booster	No	No	No
práctica-mente	No	No	No	No	No	No	Truth- Booster	No	No	No
precisa-mente	No	No	No	No	No	No	Booster Truth-	No	No	No
presunta-mente	No	No	No	No	No	No	Booster Truth-	No	No	No
pues es lo que te digo	Col	No	No	No	Consequence	No	Booster Truth-	No	No	No
pues eso mismo	Col	No	No	No	No	No	Booster Truth-	No	No	No
pues sí que	Col	No	No	No	No	No	Booster	No	No	No

Tabla 5.18: Marcadores de modalización: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
qué duda cabe queda claro ?	Col	No	No	No	No	No	Truth-Booster	No	No	No
qué duda cabe queda claro todo esto ?	Col	Sí	No	No	No	No	Truth-Booster	No	No	Social Interlocutor
realmente	No	No	No	No	No	No	Truth-Booster	No	No	Social Interlocutor
se ve que	Col	Sí	Sight	Knowledge	No	No	Truth-Booster	Inference	No	No
seguramente	No	No	No	No	No	No	Truth-Booster	No	No	No
seguramente	Col	No	No	No	No	No	Truth-Booster	No	No	No
sí	Col	No	No	No	No	No	Truth-Booster	No	No	No
seguro que	Col	No	No	No	No	No	Truth-Booster	No	No	No
si bien es verdad que	Col	Sí	No	No	Concessiveness	No	Truth-Booster	No	No	No
sí es cierto que	Col	No	No	No	No	No	Truth-Booster	No	No	No
si es que es verdad	Col	No	No	No	Coargumentation3	No	Truth-Booster	No	No	No
si es que es eso	Col	No	No	No	Coargumentation3	No	Truth-Booster	No	No	Textual Contextual
sí es verdad	Col	No	No	No	No	No	Truth-Booster	No	No	No
sí que es cierto	Col	No	No	No	No	No	Truth-Booster	No	No	No
sí que es verdad	Col	No	No	No	No	No	Truth-Booster	No	No	No
si te digo la verdad	Col	No	No	No	No	No	Truth-Booster	No	No	No
verdad	Col	No	No	No	No	No	Truth-Booster	No	No	No

Tabla 5.18: Marcadores de modalización: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
siempre	Col	No	No	No	No	No	Truth-Booster	Inference	No	No
siempre que	Col	No	No	No	No	No	Truth-Booster	Inference	No	No
sin duda	Col	No	No	No	No	No	Truth-Booster	No	No	No
también es verdad	Col	No	No	No	No	No	Truth-Booster	No	No	No
ya ves	Loc	Sí	Sight	Argumentation	No	No	Truth-Booster	No	No	SocialInter-locutor
ya ves que	Loc	Sí	Sight	Argumentation	No	No	Truth-Booster	No	No	SocialInter-locutor
justo	No	No	No	No	No	No	Truth-Booster	No	No	No
digo	No	No	No	No	No	No	Booster	No	No	No
así que por eso te digo	Col	No	No	No	Consequence	No	DRBooster	No	No	Textual-Contextual
así que ya te digo	Col	No	No	No	Consequence	No	DRBooster	No	No	No
es que no veas	Col	No	No	No	Quantity	No	DRBooster	No	No	No
no veas	Loc	Sí	No	No	Quantity	No	DRBooster	No	No	No
que no veas	Loc	Sí	Sight	Quantity	Quantity	Negative-Evaluation	DRBooster	No	No	No
que te cagas	Loc	Sí	Human-Body	Quantity	Quantity	No	DRBooster	No	No	No
justo	No	No	No	No	Time	No	DRBooster	No	No	No
la hostia	Loc	Sí	Religion	Intensity	No	Positive-Evaluation	Emotion-Booster	No	No	No
¡ole!	No	No	No	No	No	Positive-Evaluation	Emotion-Booster	No	Compliment	No

Tabla 5.18: Marcadores de modalización: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
bien es	Col	No	No	No	No	Positive-Evaluation	Emotion-Booster	No	No	No
cierto bien es	Col	No	No	No	No	Positive-Evaluation	Emotion-Booster	No	No	No
cierto que esto ya es	Col	Sí	No	No	No	Surprise	Emotion-Booster	No	No	No
muy fuerte la mar de bien	Loc	No	No	No	No	Positive-Evaluation	Emotion-Booster	No	No	No
¡coño!	No	Sí	Human-Body	Emotion	No	Surprise	Emotion-Booster	No	No	No
¡coño!	No	Sí	Human-Body	Emotion	No	Negative-Evaluation	Emotion-Booster	No	No	No
coño	No	Sí	Human-Body	Emotion	No	Surprise	Emotion-Booster	No	No	No
coño	No	Sí	Human-Body	Emotion	No	Negative-Evaluation	Emotion-Booster	No	No	No
¡ajá!	No	No	No	No	No	No	Acte-Booster	No	Assent	No
de eso nada	Loc	No	No	No	No	No	Acte-Booster	No	Ask	SocialInter-locutor
faltaría más	Loc	No	No	No	No	No	Acte-Booster	No	Acceptation	No
nanai	No	No	No	No	No	No	Acte-Booster	No	Ask	No
naranjas de la China	Loc	No	No	No	No	No	Acte-Booster	No	Ask	No
ni de coña	Loc	No	No	No	No	No	Acte-Booster	No	Ask	No
¡madre del amor hermoso!	Loc	Sí	Family	Surprise	No	Surprise	Surprise-Booster	No	No	No
¡madre mía!	Loc	Sí	Family	Surprise	No	Surprise	Surprise-Booster	No	No	No

Tabla 5.18: Marcadores de modalización: Truth booster, DR Booster, Emotion Booster, Acte Booster y Surprise Booster

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
de la hostia	Loc	Sí	Religion	Intensity	No	Surprise	Surprise- Booster	No	No	No
madre mía	Loc	Sí	No	No	No	Surprise	Surprise- Booster	No	No	No
de mi vida	Col	No	No	No	No	Surprise	Surprise- Booster	No	No	No
no me lo	Col	Sí	No	No	No	Surprise	Surprise- Booster	No	No	No
puedo creer	Col	Sí	Agreement	No	No	No	Truth	No	No	No
qué barbaridad	Loc	Sí	Agreement	No	No	No	Booster	No	No	No
ya te digo	Loc	Sí	Agreement	No	No	No	Booster	No	No	No

Tabla 5.18: Marcadores de modalización: Truth booster, DR
Booster, Emotion Booster, Acte Booster y Surprise Booster

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
bueno	No	Sí	Goodness	Discourse	No	No	Truth-Hedge	No	No	No
digamos	No	No	No	No	Re-formulation	No	Truth-Hedge	No	No	No
lo que se dice	Col	No	No	No	No	No	Truth-Hedge	Others	No	No
por decir algo	Col	No	No	No	No	No	Truth-Hedge	No	No	No
que digamos	Col	No	No	No	No	No	Truth-Hedge	No	No	No
quiero decir	Col	No	No	No	Re-formulation	No	Truth-Hedge	No	No	No
vamos	No	No	No	No	Re-formulation	No	Truth-Hedge	No	No	No
¡bah!	No	No	No	No	No	Negative-Evaluation	Truth-Hedge	No	No	No
¡buah!	No	No	No	No	No	Negative-Evaluation	Truth-Hedge	No	No	No
a mi modo de ver	Col	No	No	No	No	No	Truth-Hedge	No	No	No
aproximadamente	No	No	No	No	No	No	Truth-Hedge	No	No	No
bah	No	No	No	No	No	Negative-Evaluation	Truth-Hedge	No	No	No

Tabla 5.19: Marcadores de modalización: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
como que	Loc	No	No	No	No	No	Truth-Hedge	No	No	No
desde mi punto de vista	Col	Sí	Sight	Opinion	Fra-me-work	No	Truth-Hedge	First-Person	No	No
desde nuestro punto de vista	Col	Sí	Sight	Opinion	Fra-me-work	No	Truth-Hedge	First-Person	No	No
en mi opinión	Col	No	No	No	Fra-me-work	No	Truth-Hedge	First-Person	No	No
en parte	Loc	Sí	Space	The-Truth	Coargu-men-tation2	No	Truth-Hedge	No	No	No
entre comillas	Loc	Sí	No	No	No	No	Truth-Hedge	No	No	No
entre paréntesis	Loc	Sí	No	No	No	No	Truth-Hedge	No	No	No
es casi seguro que	Col	No	No	No	No	No	Truth-Hedge	No	No	No
más bien	Loc	No	No	No	No	No	Truth-Hedge	No	No	No
más bien al contrario	Col	No	No	No	Con-tra-argu-men-tation	No	Truth-Hedge	No	No	No
más bien que	Loc	No	No	No	No	Positive-Evaluation	Truth-Hedge	No	No	No
más o menos	Col	No	No	No	No	No	Truth-Hedge	No	No	No

Tabla 5.19: Marcadores de modalización: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
miento	No	No	No	No	Re- for- mula- tion	No	Truth- Hedge	No	No	No
o algo	Col	No	No	No	Inde- finite- Transi- tional- End	No	Truth- Hedge	No	No	No
o algo ?	Col	No	No	No	No	No	Truth- Hedge	No	No	So- cial- Inter- locu- tor
o algo de eso	Col	No	No	No	Inde- finite- Transi- tional- End	No	Truth- Hedge	No	No	No
o algo por el esti- lo	Col	No	No	No	Inde- finite- Transi- tional- End	No	Truth- Hedge	No	No	No
o cómo ?	Col	No	No	No	No	No	Truth- Hedge	No	No	So- cial- Inter- locu- tor

Tabla 5.19: Marcadores de modalización: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
o no ?	Col	No	No	No	Option	No	Truth-Hedge	No	No	Social-Interactional
no ?	Col	No	No	No	Option	No	Truth-Hedge	No	No	Social-Interactional
o qué ?	Col	No	No	No	Option	No	Truth-Hedge	No	No	Social-Interactional
o sí ?	Col	No	No	No	Option	No	Truth-Hedge	No	No	Social-Interactional
por así decirlo	Col	No	No	No	Re-formulation	No	Truth-Hedge	No	No	No
por lo visto	Loc	No	No	No	No	No	Truth-Hedge	Others	No	No
posiblemente	No	No	No	No	No	No	Truth-Hedge	No	No	No
probablemente	No	No	No	No	No	No	Truth-Hedge	No	No	No

Tabla 5.19: Marcadores de modalización: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
que a saber	Col	No	No	No	No	No	Truth-Hedge	Others	No	No
qué coño mal	Col	Sí	No	No	No	Negative-Evaluation	Truth-Hedge	No	No	No
qué digo	Col	No	No	No	Re-formulation	No	Truth-Hedge	No	No	No
que recuerde	Col	No	No	No	No	No	Truth-Hedge	No	No	No
qué te vas a acordar	Col	No	No	No	No	Negative-Evaluation	Truth-Hedge	No	No	Social-Interlocutor
que yo sepa	Col	No	No	No	No	No	Truth-Hedge	No	No	No
quiero decir que	Col	No	No	No	Re-formulation	No	Truth-Hedge	No	No	No
quizá	No	No	No	No	No	No	Truth-Hedge	No	No	No
quizás	No	No	No	No	No	No	Truth-Hedge	No	No	No
relativamente	No	No	No	No	No	No	Truth-Hedge	No	No	No
se supone	Col	No	No	No	No	No	Truth-Hedge	Others	No	No
se supone que	Col	No	No	No	No	No	Truth-Hedge	Others	No	No

Tabla 5.19: Marcadores de modalización: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
según parece	Col	No	No	No	No	No	Truth-Hedge	Others	No	No
si	No	No	No	No	Hy-po-thesis	No	Truth-Hedge	No	Assertion	No
si acaso	Col	No	No	No	Hy-po-thesis	No	Truth-Hedge	No	Assertion	No
supongo	Col	No	No	No	No	No	Truth-Hedge	First-Person	No	No
supongo que	Col	No	No	No	No	No	Truth-Hedge	First-Person	No	No
supongo que sí	Col	No	No	No	No	No	Truth-Hedge	First-Person	No	No
supongo yo	Col	No	No	No	No	No	Truth-Hedge	First-Person	No	No
supuestamente	No	No	No	No	No	No	Truth-Hedge	First-Person	No	No
tal vez	Loc	No	No	No	No	No	Truth-Hedge	No	No	No
vete a saber	Loc	Sí	No	No	No	No	Truth-Hedge	No	No	No
vete tú a saber	Loc	Sí	No	No	No	No	Truth-Hedge	No	No	No
vete tú hhh a saber	Loc	Sí	No	No	No	No	Truth-Hedge	No	No	No
ya ves tú	Loc	Sí	Sight	Argumentation	No	No	Truth-Hedge	No	No	No

Tabla 5.19: Marcadores de modalización: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
¡hombre!	No	Sí	Human- Being	Hedge	No	No	Truth- Hedge	No	No	Social- Inter- locu- tor
hombre	No	Sí	Human- Being	Hedge	No	No	Truth- Hedge	No	No	Social- Inter- locu- tor
¡jobar!	No	No	No	No	No	Negative- Evaluation	Emo- tion- Hedge	No	No	No
¡jolín!	No	No	No	No	No	Negative- Evaluation	Emo- tion- Hedge	No	No	No
¡jelines!	No	No	No	No	No	Negative- Evaluation	Emo- tion- Hedge	No	No	No
¡jope!	No	No	No	No	No	Negative- Evaluation	Emo- tion- Hedge	No	No	No
¡leche!	No	Sí	Food	Com- plaint	No	Negative- Evaluation	Emo- tion- Hedge	No	No	No
jolín	No	No	No	No	No	Negative- Evaluation	Emo- tion- Hedge	No	No	No

Tabla 5.19: Marcadores de modalización: TruthHedge, Emo-
tion Hedge, Hedge Action, DRHedge, InteractionalHedge,
ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
cariño	No	No	No	No	No	No	- Hedge	No	No	Social-Interactional
a lo mejor	Col	No	No	No	Hypothesis	Positive-Evaluation	DR-Hedge	No	No	No
a lo peor	Col	No	No	No	Hypothesis	Negative-Evaluation	DR-Hedge	No	No	No
a menos que	Loc	No	No	No	Hypothesis	No	DR-Hedge	No	No	No
a no ser que	Loc	No	No	No	Hypothesis	No	DR-Hedge	No	No	No
así que bueno	Col	No	No	No	Consequence	No	DR-Hedge	No	No	No
así que nada	Col	No	No	No	Consequence	No	DR-Hedge	No	No	No
así que no sé	Col	No	No	No	Consequence	No	DR-Hedge	No	No	No
como si	Col	No	No	No	Hypothesis	No	DR-Hedge	No	Assertion	No

Tabla 5.19: Marcadores de modalización: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
venga	No	No	No	No	No	No	Interactional-Hedge	Interactional-No	Interactional-Close	Social-Interactional-locutor
venga	No	No	No	No	No	No	Interactional-Hedge	Interactional-No	Action	Social-Interactional-locutor
perdón	No	No	No	No	No	No	Acte-Hedge	No	Apolo-gize	Social-Interactional-locutor
qué demonios	Loc	Sí	Mi-thology	- Hedge	No	No	Acte-Hedge	No	Risk	No
hija mía	Col	No	No	No	No	No	Acte-Hedge	No	No	Social-Interactional-locutor
no por otra cosa	Col	No	No	No	No	No	Acte-Hedge	No	No	No
no te digo nada	Col	No	No	No	No	No	Acte-Hedge	No	No	Social-Interactional-locutor

Tabla 5.19: Marcadores de modalización: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
no te digo que no	Col	No	No	No	No	No	Acte-Hedge	No	No	No
no te jode	Loc	Sí	No	No	No	Negative-Evaluation	Acte-Hedge	No	Rejection	No
por favor	Loc	No	No	No	No	No	Acte-Hedge	No	No	Social-Interactional
porfa	Loc	No	No	No	No	No	Acte-Hedge	No	No	Social-Interactional
nada	No	No	No	No	Indefinite-Transitional	No	Formulation-Hedge	No	No	No
no sé cuántos	Col	No	No	No	Indefinite-Transitional	No	Formulation-Hedge	No	No	No
no sé qué no sé cuántos	Col	No	No	No	Indefinite-Transitional	No	Formulation-Hedge	No	No	No

Tabla 5.19: Marcadores de modalización: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionalHedge, ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
y eso	Col	No	No	No	Indefinite-Transitional-End	No	FormulationHedge	No	No	Textual-Contextual
y nada	Col	No	No	No	Indefinite-Transitional-End	No	FormulationHedge	No	No	Textual-Contextual
y todo	Col	No	No	No	Indefinite-Transitional-End	No	FormulationHedge	No	No	Textual-Contextual
y ya	Col	No	No	No	Indefinite-Transitional-End	No	FormulationHedge	No	No	No

Tabla 5.19: Marcadores de modalización: TruthHedge, Emotion Hedge, Hedge Action, DRHedge, InteractionHedge, ActeHedge, FormulationHedge

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
escucha	No	No	No	No	No	No	Interaction	No	No	Social
mira	No	No	No	No	No	No	Interaction	No	No	Interlocutor Social
ojo	No	Sí	Body	Warning	No	No	Interaction	No	Warning	Interlocutor Social
verdad ?	No	No	No	No	No	No	Interaction	No	No	Interlocutor Social
ya	No	No	No	No	No	No	Interaction	No	No	Interlocutor No
¡ah! sí ?	Col	No	No	No	No	No	Interaction	No	No	No
¡eh!	No	No	No	No	No	No	Interaction	No	No	Social
¡ey!	No	No	No	No	No	No	Interaction	No	No	Interlocutor Social
a que sí ?	Loc	No	No	No	No	No	Interaction	No	No	Interlocutor Social
ah ?	No	No	No	No	No	No	Interaction	No	No	Interlocutor Social
cómo ?	No	No	No	No	No	No	Interaction	No	No	Interlocutor Social
cómo lo ves ?	Col	Sí	Sight	Opinion	No	No	Interaction	No	No	Interlocutor Social
cómo que no ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor Social
de verdad ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor Social
eh ?	No	No	No	No	No	No	Interaction	No	No	Interlocutor Social
entiendes ?	No	No	No	No	No	No	Interaction	No	No	Interlocutor Social
entonces ?	No	No	No	No	Consequence	No	Interaction	No	No	Interlocutor Social
es o no es cierto esto ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor Social

Tabla 5.20: Marcadores de modalización: Interaction

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
lo ves ?	Col	Sí	No	No	No	No	Interaction	No	No	Social
me entiendes ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor
me entiendes lo que te quiero decir ?	Col	No	No	No	No	No	Interaction	No	No	Social
no te acuerdas ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor
no ves ?	Col	Sí	No	No	No	No	Interaction	No	No	Social
os acordáis ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor
perdona ?	No	No	No	No	No	No	Interaction	No	No	Social
qué creéis ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor
qué dices ?	Col	No	No	No	No	No	Interaction	No	No	Social
que no ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor
qué pasa ?	Col	No	No	No	No	No	Interaction	No	No	Social
qué te iba a decir ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor
sabes ?	No	No	No	No	No	No	Interaction	No	No	Social
sabes una cosa ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor
seguro ?	No	No	No	No	No	No	Interaction	No	No	Social
sí ?	No	No	No	No	No	No	Interaction	No	No	Interlocutor

Tabla 5.20: Marcadores de modalización: Interaction

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
sí o no ?	Col	No	No	No	No	No	Interaction	No	No	Social
te acuerdas ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor Social
vale ?	No	Sí	No	No	No	No	Interaction	No	No	Interlocutor Social
verdad hijo ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor Social
verdad que no ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor Social
ves ?	No	Sí	No	No	No	No	Interaction	No	No	Interlocutor Social
ves cómo es cierto ?	Col	Sí	No	No	No	No	Interaction	No	No	Interlocutor Social
y qué pasa ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor Social
y ya está ?	Col	No	No	No	No	No	Interaction	No	No	Interlocutor Social
igualmente	No	No	No	No	No	No	Interaction	No	Reciprocal-Action	No

Tabla 5.20: Marcadores de modalización: Interaction

5.4.4. Marcadores de Evidencialidad

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
como es natural	Col	No	No	No	No	No	Truth Booster	Cultural Knowledge	No	No
eso es como todo	Col	No	No	No	Generaliza- tion	No	No	Inference	No	No
lógicamente	No	No	No	No	No	No	Truth Booster	Inference	No	No
normalmente	No	Sí	No	No	Generaliza- tion	No	No	Inference	No	No
por fuerza	Loc	No	No	No	No	No	Truth Booster	Inference	No	No
por huevos	Loc	No	No	No	No	No	Truth Booster	Inference	No	No
por lo general	Loc	No	No	No	Generaliza- tion	No	No	Inference	No	No
se ve que	Col	Sí	Sight	Knowledge	No	No	Truth Booster	Inference	No	No
siempre	Col	No	No	No	No	No	Truth Booster	Inference	No	No
siempre que	Col	No	No	No	No	No	Truth Booster	Inference	No	No
dicen que	Col	No	No	No	No	No	No	Others	No	No
lo que se dice	Col	No	No	No	No	No	Truth Hedge	Others	No	No
por lo visto	Loc	No	No	No	No	No	Truth Hedge	Others	No	No
que a saber	Col	No	No	No	No	No	Truth Hedge	Others	No	No
se supone	Col	No	No	No	No	No	Truth Hedge	Others	No	No
se supone que	Col	No	No	No	No	No	Truth Hedge	Others	No	No
según parece	Col	No	No	No	No	No	Truth Hedge	Others	No	No

Tabla 5.21: Marcadores de evidencialidad: Cultural Knowledge, inference, others, first person

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
lo que te digo	Col	No	No	No	No	No	Truth	First Person	No	No
desde mi punto de vista	Col	Sí	Sight	Opinion	Framework	No	Booster Truth	First Person	No	No
desde nuestro punto de vista	Col	Sí	Sight	Opinion	Framework	No	Hedge Truth	First Person	No	No
en mi opinión	Col	No	No	No	Framework	No	Truth Hedge	First Person	No	No
lo que te digo	Col	No	No	No	No	No	Truth	First Person	No	No
supongo	Col	No	No	No	No	No	Booster Truth	First Person	No	No
supongo que	Col	No	No	No	No	No	Hedge Truth	First Person	No	No
supongo que sí	Col	No	No	No	No	No	Truth Hedge	First Person	No	No
supongo yo	Col	No	No	No	No	No	Truth Hedge	First Person	No	No
supuestamente	No	No	No	No	No	No	Truth Hedge	First Person	No	No
digo	No	No	No	No	No	No	No	First Person	No	No

Tabla 5.21: Marcadores de evidencialidad: Cultural Knowledge, inference, others, first person

5.4.5. Marcadores de Actos de Habla

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
verás	No	Sí	No	No	No	No	No	No	Commitment	Social Interlocutor
que en paz descanse	Col	No	No	No	No	No	No	No	Wish	No
no me jodas	Loc	Sí	No	No	No	Negative Evaluation	No	No	Rejection	No
no te jode	Loc	Sí	No	No	No	Negative Evaluation	ActeHedge	No	Rejection	No
que le den por culo	Loc	Sí	Body	Rejection	No	Negative Evaluation	No	No	Rejection	No
faltaría más	Loc	No	No	No	No	No	Acte	No	Acceptation	No
de eso nada	Loc	No	No	No	No	No	Booster Acte	No	Ask	Social Interlocutor
nanai	No	No	No	No	No	No	Booster Acte	No	Ask	No
naranjas de la China	Loc	No	No	No	No	No	Booster Acte	No	Ask	No
ni de coña	Loc	No	No	No	No	No	Booster Acte	No	Ask	No
como si	Col	No	No	No	Hypothesis	No	Booster DR Hedge	No	Assertion	No
si	No	No	No	No	Hypothesis	No	Truth Hedge	No	Assertion	No
si acaso	Col	No	No	No	Hypothesis	No	Truth Hedge	No	Assertion	No
¡ole!	No	No	No	No	No	Positive Evaluation	Emo- tion- Booster Acte	No	Compliment	No
¡ajá!	No	No	No	No	No	No	Booster	No	Assent	No
igualmente	No	No	No	No	No	No	Interaction	No	Reciprocal- Action	No
venga	No	No	No	No	No	No	Interactional Hedge	No	Interaction- Close	Social Interlocutor

Tabla 5.22: Marcadores de actos de habla: Commitment,

wish, rejection, acceptance, ask, compliment, assent, recip-

rocal action, interaction close, risk, justification, complaint,

resignation, greeting open, refuse, greeting close, gratitude.

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
qué demonios	Loc	Sí	Mi- tho- logy	Hedge	No	No	Acte- Hedge	No	Risk	No
es que	Loc	Sí	Exis- tence	Explanation	Explanation	No	No	No	Justification	No
el caso es que	Col	No	No	No	Concretness	No	No	No	Justification	No
la cosa es que	Col	No	No	No	Concretness	No	No	No	Justification	No
la cuestión es que	Col	No	No	No	Concretness	No	No	No	Justification	No
lo que pasa es que	Col	No	No	No	Concretness	No	No	No	Justification	No
lo que pasa que	Col	No	No	No	Concretness	No	No	No	Justification	No
hay que ver	Loc	Sí	Sight	Complaint	No	No	No	No	Complaint	No
que si	Col	No	No	No	Coar- gu- men- ta- tion2	Negative Evaluation	No	No	Complaint	No
en fin	Loc	Sí	End	Discourse	Conclusion Summary	No	No	No	Resignation	No
qué tal	Loc	No	No	No	No	No	No	No	Greeting Open	No
buenos días	Loc	No	No	No	No	No	No	No	Greeting Open	No
buenos días a todos	Loc	No	No	No	No	No	No	No	Greeting Open	No
hola	No	No	No	No	No	No	No	No	Greeting Open	No

Tabla 5.22: Marcadores de actos de habla: Commitment, wish, rejection, acceptance, ask, compliment, assent, reciprocal action, interaction close, risk, justification, complaint, resignation, greeting open, refuse, greeting close, gratitude, action, warning

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
muy buenos días	Col	No	No	No	No	No	No	No	Greeting Open	No
no te digo	Loc	Sí	Action	Rejection	No	Negative Evaluation	Truth Booster	No	Refuse	No
una mierda	No	No	No	No	No	Negative Evaluation	No	No	Refuse	No
adiós	No	No	No	No	No	No	No	No	Greeting	Social Interlocutor
buenas noches	Loc	No	No	No	No	No	No	No	Close Greeting	No
hasta luego	Loc	No	No	No	No	No	No	No	Close Greeting	No
muy buenas noches	Col	No	No	No	No	No	No	No	Close Greeting	No
gracias	No	No	No	No	No	No	No	No	Close Gratitude	No
muchas gracias	Loc	No	No	No	No	No	No	No	Gratitude	No
adelante	No	No	No	No	No	No	No	No	Action	Social Interlocutor
anda	No	Sí	Action	Interaction	No	No	No	No	Action	Social Interlocutor
ya sabes	Col	No	No	No	No	No	No	No	Action	No
¡hala!	No	No	No	No	No	Negative Evaluation	No	No	Action	No
hala	No	No	No	No	No	Negative Evaluation	No	No	Action	No
venga	No	No	No	No	No	No	Interactional-Hedge	No	Action	Social Interlocutor
atención	No	No	A-tention	Warning	No	No	No	No	Warning	Social-Interlocutor
cuidado	No	No	No	No	No	No	No	No	Warning	Social Interlocutor

Tabla 5.22: Marcadores de actos de habla: Commitment, wish, rejection, acceptance, ask, compliment, assent, reciprocal action, interaction close, risk, justification, complaint, resignation, greeting open, refuse, greeting close, gratitude,

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
ojo	No	Sí	Body	Warning	No	No	Interaction	No	Warning	Social
¡ojo!	No	Sí	Body	Warning	No	No	No	No	Warning	Interlocutor
tú verás	Col	No	No	No	No	No	No	No	Warning	No
ya lo verás	Col	Sí	Sight	Future	No	No	No	No	Warning	Social
ya verás	Loc	Sí	Sight	Future	No	No	No	No	Warning	Interlocutor
ya verás tú	Loc	Sí	Sight	Future	No	No	No	No	Warning	Social
										Interlocutor

Tabla 5.22: Marcadores de actos de habla: Commitment, wish, rejection, acceptance, ask, compliment, assent, reciprocal action, interaction close, risk, justification, complaint, resignation, greeting open, refuse, greeting close, gratitude, action, warning

5.4.6. Marcadores de Deixis

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
a ver	Loc	Sí	Sight	De- monstra- tion	Focus	No	No	No	No	SocialInterlocutor
adelante	No	No	No	No	No	No	No	No	Ac- tion	SocialInterlocutor
anda	No	Sí	Ac- tion A-	Interac- tion	No	No	No	No	Ac- tion	SocialInterlocutor
atención	No	No	atten- tion	War- ning	No	No	No	No	War- ning	SocialInterlocutor
cariño	No	No	No	No	No	No	Hedge- Ac- tion	No	No	SocialInterlocutor
cuidado	No	No	No	No	No	No	No	No	War- ning	SocialInterlocutor
escucha	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
eso	No	No	No	No	No	No	Truth- Booster	No	No	SocialInterlocutor
hija	No	No	No	No	No	No	No	No	No	SocialInterlocutor
macho	No	No	No	No	No	No	No	No	No	SocialInterlocutor
maja	No	No	No	No	No	No	No	No	No	SocialInterlocutor
majo	No	No	No	No	No	No	No	No	No	SocialInterlocutor
mira	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
mujer	No	No	No	No	No	No	No	No	No	SocialInterlocutor
ojo	No	Sí	Body	War- ning	No	No	Interac- tion	No	War- ning	SocialInterlocutor
perdón	No	No	No	No	No	No	Acte- Hedge	No	A- polo- gize	SocialInterlocutor

Tabla 5.23: Marcadores de deixis: Social Interlocutor

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
señor	No	No	No	No	No	No	No	No	No	SocialInterlocutor
tía	No	No	No	No	No	No	No	No	No	SocialInterlocutor
tío	No	No	No	No	No	No	No	No	No	SocialInterlocutor
vamos a ver	Loc	Sí	Sight	De- monstra- tion	Dis- course- Produc- tion	No	No	No	No	SocialInterlocutor
verdad ?	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
usted	No	No	No	No	No	No	No	No	No	SocialInterlocutor
adiós	No	No	No	No	No	No	No	No	Greeting- Close	SocialInterlocutor
¡eh!	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
¡ey!	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
a que sí ?	Loc	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
ah ?	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
así que mira	Col	No	No	No	Con- se- quence	No	No	No	No	SocialInterlocutor
cómo ?	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
cómo lo ves ?	Col	Sí	Sight	Opinion	No	No	Interac- tion	No	No	SocialInterlocutor
cómo que no ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
de eso nada	Loc	No	No	No	No	No	Acte- Booster	No	Ask	SocialInterlocutor
de verdad ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor

Tabla 5.23: Marcadores de deixis: Social Interlocutor

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
eh ?	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
entiendes ?	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
entonces ?	No	No	No	No	Con- se- quence	No	Interac- tion	No	No	SocialInterlocutor
es o no es cierto esto ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
eso mismo	Col	No	No	No	No	No	Truth- Booster	No	No	SocialInterlocutor
fíjate	No	No	No	No	No	No	No	No	No	SocialInterlocutor
hija mía	Col	No	No	No	No	No	Acte- Hedge	No	No	SocialInterlocutor
lo ves ?	Col	Sí	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
me entiendes ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
me entiendes lo que te quiero de- cir ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
no te acuerdas ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
no te digo nada	Col	No	No	No	No	No	Acte- Hedge	No	No	SocialInterlocutor
no ves ?	Col	Sí	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
o algo ?	Col	No	No	No	No	No	Truth- Hedge	No	No	SocialInterlocutor
o cómo ?	Col	No	No	No	No	No	Truth- Hedge	No	No	SocialInterlocutor
o no ?	Col	No	No	No	Op- tion	No	Truth- Hedge	No	No	SocialInterlocutor

Tabla 5.23: Marcadores de deixis: Social Interlocutor

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
no ?	Col	No	No	No	Op- tion	No	Truth- Hedge	No	No	SocialInterlocutor
o qué ?	Col	No	No	No	Op- tion	No	Truth- Hedge	No	No	SocialInterlocutor
o sí ?	Col	No	No	No	Op- tion	No	Truth- Hedge	No	No	SocialInterlocutor
oíga	No	No	No	No	No	No	No	No	No	SocialInterlocutor
os acordáis ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
oye	No	No	No	No	No	No	No	No	No	SocialInterlocutor
perdona	No	No	No	No	No	No	No	No	A- polo- gize	SocialInterlocutor
perdona ?	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
perdóname	No	No	No	No	No	No	No	No	A- polo- gize	SocialInterlocutor
perdóname un momento	Col	No	No	No	No	No	No	No	A- polo- gize	SocialInterlocutor
por favor	Loc	No	No	No	No	No	Acte- Hedge	No	No	SocialInterlocutor
porfa	Loc	No	No	No	No	No	Acte- Hedge	No	No	SocialInterlocutor
qué creéis ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
qué dices ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
que no ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor

Tabla 5.23: Marcadores de deixis: Social Interlocutor

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
qué pasa ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
qué te iba a decir ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
qué te vas a acordar	Col	No	No	No	No	Negative- E- valuation	Truth- Hedge	No	No	SocialInterlocutor
queda claro ?	Col	Sí	No	No	No	No	Truth- Booster	No	No	SocialInterlocutor
queda claro todo esto ?	Col	Sí	No	No	No	No	Truth- Booster	No	No	SocialInterlocutor
sabes ?	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
sabes una cosa ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
seguro ?	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
sí ?	No	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
sí o no ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
te acuerdas ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
tía petarda	Col	Sí	No	No	No	No	No	No	No	SocialInterlocutor
vale ?	No	Sí	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
verás	No	Sí	No	No	No	No	No	No	Com- mit- ment	SocialInterlocutor
verdad hijo ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor

Tabla 5.23: Marcadores de deixis: Social Interlocutor

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
verdad que no ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
ves ?	No	Sí	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
ves cómo es cierto ?	Col	Sí	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
y qué pasa ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
y ya está ?	Col	No	No	No	No	No	Interac- tion	No	No	SocialInterlocutor
ya lo verás	Col	Sí	Sight	Future	No	No	No	No	Warning	SocialInterlocutor
ya verás	Loc	Sí	Sight	Future	No	No	No	No	Warning	SocialInterlocutor
ya verás tú	Loc	Sí	Sight	Future	No	No	No	No	Warning	SocialInterlocutor
ya ves	Loc	Sí	Sight	Argumenta- tion	No	No	Truth- Booster	No	No	SocialInterlocutor
ya ves que	Loc	Sí	Sight	Argumenta- tion	No	No	Truth- Booster	No	No	SocialInterlocutor
¡hombre!	No	Sí	Human- Being	Surpri- se	No	Surpri- se	No	No	No	SocialInterlocutor
¡hombre!	No	Sí	Human- Being	Hedge	No	No	Truth- Hedge	No	No	SocialInterlocutor
hombre	No	Sí	Human- Being	Surpri- se	No	Surpri- se	No	No	No	SocialInterlocutor
hombre	No	Sí	Human- Being	Hedge	No	No	Truth- Hedge	No	No	SocialInterlocutor

Tabla 5.23: Marcadores de deixis: Social Interlocutor

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
venga	No	No	No	No	No	No	Interac- tional- Hedge	No	Interac- tion- Close	SocialInterlocutor
venga	No	No	No	No	No	No	Interac- tional- Hedge	No	Ac- tion	SocialInterlocutor

Tabla 5.23: Marcadores de deixis: Social Interlocutor

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
anteriormente	No	Sí	Time	Dis- course	Time	No	No	No	No	TextualContextual
aparte de	Loc	Sí	Space	Dis- course	Coar- gumen- ta- tion2	No	No	No	No	TextualContextual
aparte	No	Sí	Space	Dis- course	Coar- gumen- ta- tion2	No	No	No	No	TextualContextual
así es	Col	No	No	No	No	No	Truth- Booster	No	No	TextualContextual
de ahí	Loc	Sí	Space	Dis- course	Ana- phoric- Cause- Cata- phoric- Conse- quence	No	No	No	No	TextualContextual
de aquí	Loc	Sí	Space	Dis- course	Ana- phoric- Cause- Cata- phoric- Conse- quence	No	No	No	No	TextualContextual

Tabla 5.24: Marcadores de deixis: Textual Contextual

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
de entrada	Loc	Sí	Space	Dis- cour- se	Coar- gumen- ta- tion1	No	No	No	No	TextualContextual
de un lado	Loc	Sí	Space	Dis- cour- se	Coar- gumen- ta- tion2	No	No	No	No	TextualContextual
esto es	Col	No	No	No	Re- for- mula- tion	No	No	No	No	TextualContextual
por ello	Col	No	No	No	Ana- phoric- Cause- Cata- phoric- Conse- quence	No	No	No	No	TextualContextual
por eso	Col	No	No	No	Ana- phoric- Cause- Cata- phoric- Conse- quence	No	No	No	No	TextualContextual
por otro	Loc	Sí	Space	Dis- cour- se	Coar- gumen- ta- tion2	No	No	No	No	TextualContextual

Tabla 5.24: Marcadores de deixis: Textual Contextual

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
por un lado	Loc	Sí	Space	Dis- course	Coar- gumen- ta- tion2	No	No	No	No	TextualContextual
primero	No	No	No	No	Coar- gumen- ta- tion1	No	No	No	No	TextualContextual
pues eso	Loc	No	No	No	No	No	Truth- Booster	No	No	TextualContextual
segundo	No	No	No	No	Coar- gumen- ta- tion2	No	No	No	No	TextualContextual
tercero	No	Sí	No	No	Coar- gumen- ta- tion2	No	No	No	No	TextualContextual
y demás	Col	No	No	No	Inde- finite- Transi- tional- End	No	No	No	No	TextualContextual
y eso que	Col	No	No	No	Con- cessi- veness Inde- finite- Transi- tional- End	No	No	No	No	TextualContextual
y eso	Col	No	No	No	Inde- finite- Transi- tional- End	No	Formula- tion- Hedge	No	No	TextualContextual

Tabla 5.24: Marcadores de deixis: Textual Contextual

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
y nada	Col	No	No	No	Indefinite-Transitional-End	No	Formula-tion-Hedge	No	No	TextualContextual
y todo	Col	No	No	No	Indefinite-Transitional-End	No	Formula-tion-Hedge	No	No	TextualContextual
a este respecto	Col	No	No	No	To-picaliza-tion	No	No	No	No	TextualContextual
al respecto	Loc	No	No	No	To-picaliza-tion	No	No	No	No	TextualContextual
ante todo	Loc	Sí	Space	Dis-course	Coar-gumentation	No	No	No	No	TextualContextual
antes de nada	Loc	Sí	Space	Dis-course	Coar-gumentation	No	No	No	No	TextualContextual
antes que nada	Loc	Sí	Space	Dis-course	Coar-gumentation	No	No	No	No	TextualContextual

Tabla 5.24: Marcadores de deixis: Textual Contextual

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
aparte de eso	Loc	Sí	Space	Dis- cour- se	Coar- gumen- ta- tion2	No	No	No	No	TextualContextual
así las cosas	Col	No	No	No	Coar- gumen- ta- tion2	No	No	No	No	TextualContextual
así que eso	Col	No	No	No	Conse- quence	No	No	No	No	TextualContextual
así que por eso te digo	Col	No	No	No	Conse- quence	No	DR- Boos- ter	No	No	TextualContextual
así que todo eso	Col	No	No	No	Conse- quence- Su- mma- ry	No	No	No	No	TextualContextual
así y todo	Loc	No	No	No	Conce- ssive- ness	No	No	No	No	TextualContextual
aún así	Loc	No	No	No	Conce- ssive- ness	No	No	No	No	TextualContextual

Tabla 5.24: Marcadores de deixis: Textual Contextual

MD	UFS	MET	SD	TD	DR	ED	MOD	EVI	SA	DEX
de ahí que	Col	No	No	No	Ana-phoric Cause-Cata-phoric Consequence	No	No	No	No	TextualContextual
					Coar-gumen-tation3	No	Truth-Booster	No	No	TextualContextual
verá	No	No	No	No	No	No	No	No	No	TextualContextual
a tomar por culo	Loc	Sí	Body	Rejection	Conclusion-Summary	Negative-Evaluation	No	No	No	TextualContextual

Tabla 5.24: Marcadores de deixis: Textual Contextual

Aquí termina la clasificación de las unidades con información pragmática de C-ORAL-ROM. Para un estudioso de la Pragmática dicha clasificación no va a ser suficiente e incluso puede llegar a considerar simplista; sin embargo es bastante rica en relación a lo que se ha hecho hasta el momento en Lingüística Computacional en lo que respecta a la lengua española. La bibliografía utilizada para este capítulo y que no ha sido citada directamente en el texto se puede consultar en la página Web.

Pasemos ahora a la parte computacional de la tesis. Esta está dividida en dos capítulos. En el primero de ellos tratamos cuestiones relativas a la Lingüística de Corpus, su anotación, la presentación de las características generales de nuestro corpus y el modo en que quedan etiquetadas finalmente las conversaciones en formato XML. En el segundo capítulo vamos a explicar la metodología que hemos seguido para etiquetar el corpus.

Parte III

Del procesamiento

Capítulo 6

Lingüística de corpus y Anotación pragmática: C-ORAL-ROM y Pragmatext

6.1. Introducción

Este capítulo está dedicado a enmarcar nuestro trabajo dentro de la disciplina de Lingüística de Corpus. Hemos dividido la exposición de los conocimientos en tres grandes partes. En la primera, presentamos las características de diseño más interesantes del corpus con el que vamos a trabajar, al tiempo que lo insertamos dentro las líneas de investigación del pasado siglo en la recolección de muestras de oralidad. En la segunda parte, tratamos el tema de la anotación de corpus y nos centramos brevemente en aquellos estudios que han hecho anotación a nivel semántico-pragmático. Finalmente, en la tercera parte, presentamos al lector el aspecto de los textos del corpus C-ORAL-ROM una vez anotados con nuestro modelo de anotación PRAGMATEXT.

6.2. C-ORAL-ROM en la tradición lingüística de corpus orales

La **Lingüística de Corpus** y los **Estudios sobre la Lengua Oral** son disciplinas independientes, pero estrechamente relacionadas entre sí, que en las últimas décadas del siglo XX lograron finalmente establecerse como líneas de investigación oficiales y consolidadas a nivel institucional dentro de la Lingüística. Tal y como señala, F. Sánchez, atrás quedan los debates del pasado sobre la utilidad de los corpus para extraer principios o reglas gramaticales que caractericen la competencia lingüística. Pocos cuestionan hoy en día las

bondades de la Lingüística de Corpus sobre todo en lo que a la creación de un patrimonio lingüístico, de un tesoro cultural, se refiere (SÁNCHEZ, 1998).

Dentro de este marco, en el año 2001, el Laboratorio de Lingüística Informática comienza su participación en el proyecto de investigación europeo C-ORAL-ROM. El propósito primordial de dicho proyecto consiste en elaborar cuatro corpus en cuatro lenguas romances que posean características de diseño homogéneas, tales como el número de palabras, las situaciones comunicativas seleccionadas para la grabación o el formato de transcripción. Aparte del español, las tres lenguas restantes en las que se elaboraron los corpus fueron el portugués (realizado por la Fundación de la Universidad de Lisboa), el francés (del que se encargó la Universidad de Provenza) y el italiano (a cargo de la Universidad de Florencia).

Estos corpus poseen, respecto de los corpus orales elaborados anteriormente sobre todo en España, una serie de características que han hecho que algunos investigadores los denominaran corpus de segunda generación (MORENO, 2002b) (MORENO, 2002a). Entre dichas diferencias se pueden destacar un mayor cuidado de los aspectos legales (los participantes firman una autorización aceptando la grabación y comercialización de su voz), la calidad de las grabaciones (gracias al avance de la tecnología en este terreno) y aspectos como la representatividad (de género discursivo, en este caso). Veamos con más profundidad algunas características de este proyecto a través del corpus C-ORAL-ROM español.

6.2.1. El corpus oral del español C-ORAL-ROM: diseño y características

El corpus C-ORAL-ROM tiene aproximadamente unas 300.000 palabras. Desde un punto de vista macroestructural, en las transcripciones podemos diferenciar dos grandes bloques: la cabecera y la transcripción propiamente dicha del sonido de la grabación. A continuación explicaremos las características más importantes de cada uno de estos bloques en sendos subapartados.

6.2.2. Las cabeceras de las transcripciones

En la cabecera se recoge información de naturaleza contextual clasificada en los siguientes campos:

@Title: un breve título

@File: el nombre del archivo

@Participants: los datos sobre los participantes

1. siglas que marcarán sus intervenciones en el texto,
2. nombre por el que le llaman los otros interlocutores,
3. género (man/woman),
4. edad
 - a) A para de 18 a 25 años
 - b) B de 25 a 40 años
 - c) C de 40 a 50 años
 - d) D mayores de 60 años
5. educación
 - a) 0 para cuando no conocemos el nivel educativo, como por ejemplo, en las grabaciones de la radio
 - b) 1 hasta educación primaria
 - c) 2 hasta educación secundaria
 - d) 3 para universitarios
6. profesión
7. rol en la interacción
8. origen geográfico o posibles influencias lingüísticas

@Date: la fecha

@Place: la ciudad donde se produjo la grabación

@Situation: la situación

@Topic: el tema o subtemas que se tratan

@Source: el origen del texto

@Class: el tipo según las distinciones de informalformal indicadas en las tablas de abajo.

@Lenght: el tamaño en minutos y segundos

@Words: en número de palabras

@Acoustic quality: el nivel de la calidad del sonido, será siempre A muy buena, calidad digital o B buena.

@Transcriptor: el transcriptor responsable del texto

@Revisor: el revisor

Informal	Familiar/Private	Monologue Dialogue Conversation
	Public	Monologue Dialogue Conversation

Tabla 6.1: Distribución de la parte informal de C-Oral-Rom

Formal		
Formal in Natural Context	Media	Telephone
Political Speech	News	Private Conversation
Political Debate	Sport	Phone Call Services (ManMachine interaction)
Preaching	Interviews	
Teaching	Meteo	
Professional Explanation	Scientific Press	
Conference	Reportage	
Business	Talk shows political debate	
Law (through media)	Talk shows thematic discussions	
	Talk shows culture	
	Talk shows science	

Tabla 6.3: Distribución de la parte formal de C-Oral-Rom

Estas son las tablas de distribución de los géneros discursivos del corpus, con las cuales se ha cumplimentado el apartado de @class:

6.2.2.1. Las transcripciones

En la tabla siguiente se recogen los signos de transcripción más importantes para la comprensión de las transcripciones:

Signo	Definición	Ejemplo
* Tres mayúsculas:	Comienzo de turno. Las siglas coincidirán con la que se les ha asignado al interlocutor en la cabecera	*ABC:
*XZY:	En caso de que no se pueda identificar al interlocutor	*XZY: ¿existe el cambio climático?
%	Tras los turnos, pueden aparecer comentarios sobre el sonido o sobre la pronunciación, señalados por este signo %. Se debe tener en cuenta que la transcripción es ortográfica. Esto quiere decir que tanto los errores de producción como las expresiones que la norma no considera formas correctas se reflejan en los comentarios de tipo %alt y no en la transcripción propiamente dicha. Supongamos que un hablante dice <i>se m'a travesao un cameón</i> en lugar de <i>se me ha atravesado un camión</i> . Estas formas se colocarán en los comentarios, con números que indiquen la posición en la que se encuentra el fenómeno.	*MIG: se me ha atravesado un camión / y casi me atropella // %alt: between (2) and (3) ma %alt: (6) cameón
/	Las transcripciones están etiquetadas siguiendo la teoría de Emanuela Cresti CRESTI y MONEGLIA (2005). La marca sirve para indicar los límites de las unidades tonales (tone units).	*PAL: juro / que quiero que gane el Real Madrid // <totalmente / eh> ?
/	Cada barra / se corresponde en el sonido con un cambio en la entonación. Si aparece tras las siglas del interlocutor, indica que se continúa la unidad informativa de su anterior turno (en medio ha habido una intervención de otro participante).	*PAL: juro / que quiero que gane el Real Madrid // <totalmente / eh> ? *USE: [<] <ya // ya> // *PAL: o sea / <lo> *PRI: [<] <ya> // *PAL: / juro / total> //

Tabla 6.4: Tabla de signos de transcripción

Signo	Definición	Ejemplo
// ? ... +	Estas marcas, que significan aser- ción, interrogación, suspensión y auto/interrupción respectivamente, señalan el final de las unidades informa- tivas completas (utterances).	*PAL: <pero> ... *PRI: [<] <para> que no te pegue tu marido <una paliza> // %alt: (1) pa *PAL: [<] <y ahora que> + *NUR: o sea / Use / él / tiene un problema / y es que los partidos / Real Madrid Barcelona / y ese tipo de partidos / les &gust [/] le gusta / verlos //
&	sirve para marcar las palabras que no se han pronunciado completamente	&cantarí por <i>cantaría</i>
&	apoyos silábicos	&eh y &ah
¡!	Las interjecciones se transcriben entre signos de admiración pegados a la pa- labra. En C-ORAL-ROM no hay signos para indicar las entonación exclamativa	¡dios mío!
hhh	Se utiliza para reflejar los sonidos no lingüísticos que son relevantes para la in- formación o la entonación. Pueden cons- tituir por sí mismos un turno de palabra. En los comentarios se señala el tipo de sonido: silbido, risa, chasquelo, resoplido, etc.	*TER: no digas eso hhh // %act: (4) laugh
[/]	Existen dos marcas para indicar que el hablante, en lugar de terminar su enuncia- do, ha realizado un reinicio de la emisión. La marca [/] indica que ha vuelto a repetir lo mismo.	*PRI: <ves ? tanta carrera // tanta carrera> // *PAL: [<] <a ver listilla [\\] a ver listilla> / la que todo lo sabe // <que no pasa de ciento cincuenta mil hhh> // %act: (19) laugh

Tabla 6.4: Tabla de signos de transcripción

Signo	Definición	Ejemplo
[///]	Reformulador prosódico de tres barras, significa que ha cambiado su emisión desde un punto de vista léxico, sintáctico o semántico. El ejemplo presenta una reformulación léxica, con sintaxis intacta.	*NUR: las se ñoras son pesadas / porque / son [/] son / &eh [///] tienen mucho en cuenta el dinero //
#	La almohadilla indica una pausa larga	*JES: xxx mi pelo / como es castaño / pues si me lo decoloran un poco / pues queda más rubio // # tú que eres más moreno / pues / te tendrían que tener más tiempo / a lo mejor / puesto / para que se te quedara más +
<> y [<]	<p>También se representan los solapamientos. Los fragmentos de habla solapados aparecen entre ángulos:</p> <p><></p> <p>El último turno donde hay solapamiento siempre se comienza con este signo:</p> <p>[<]</p>	<p>*PAL: dec ía / tengo que decir que el único que pas ó de no s é cuánto / fue Use / pero bueno / es que Use / dec ía / hhh / o sea / xxx [///] <nos barrer á a todos> / %act: (22) laugh</p> <p>*NUR: [<] <qu é pelota> // qué pelota //</p> <p>Ni los fragmentos ni los hhh se han tenido en cuenta a la hora de contar el número de palabras del fichero.</p>

Tabla 6.4: Tabla de signos de transcripción

6.2.3. La anotación del corpus C-ORAL-ROM: niveles de anotación y formatos

C-ORAL-ROM consta de cuatro versiones del corpus cada una con diferentes niveles de etiquetado. La primera es en formato texto y refleja el sistema de transcripción hasta aquí mostrado. La segunda muestra en XML las marcas de alineamiento con el sonido de la grabación. Después hay una versión de C-ORAL-ROM - en formato MULTTEXT y otra en formato XML - donde el corpus está anotado con información categorial y morfosintáctica. Y por último, existe una versión del corpus también en XML anotada con información semántica, de naturaleza eventiva, para ser más exactos. A continuación, presentamos un ejemplo de conversación en el formato de transcripción original. En los anexos, se podrán consultar ejemplos de textos etiquetados en los niveles y formatos que acabamos de señalar.

@Title: problemas de novios
@File: efamcv10
@Participants: ELE, Elena, (woman, B, 3, student, participant, Madrid)
ANT, Antonio, (man, B, 3, student, participant, Madrid)
ANA, Ana, (woman, B, 3, student, participant, Madrid)
SEV, Ana, (woman, B, 3, student, participant, Sevilla)
@Date: 03/04/2001
@Place: Madrid
@Situation: chat between friends at university, hidden,
researcher participant
@Topic: friends and lovers
@Source: C-ORAL-ROM
@Class: informal, family\private, conversation
@Length: 6' 24''
@Words: 1568
@Acoustic quality: A
@Transcriber: Manuel
@Revisor: Guillermo ; Manuel and Jesús (prosody)
@Comments:

*ELE: yo le entend ía perfectamente // y lo que no le quer ía decir era /

decirle / mira / Fran / o sea / yo no estoy enamorada de ti / entonces
no lo voy a ...

*ANA: <claro> //

*ELE: [<] <no le quer ía> decir eso // entonces no [/] no [/] no + ¡ah!
/ porque adem ás es que tampoco sab ía muy bien lo que le quer ía decir //
entonces / a m í me \&pa + qu é pasa ?

*ANT: hhh / nada //

\%act: (1) laugh

*ELE: ¡ah! / entonces yo lo que hhh quer ía ...

%act: (6) laugh

*ANA: hhh //

\%act: (1) laugh

*ELE: sabes ? no s é / como que no sab ía + a m í me pasa / t ía /
que cuando no s é
qué decir / es que no digo nada / y me quedo callada // y puede
parecer que es
que me da todo igual / pero no es por eso / es porque de verdad
/ que no s é
<qué decirle> //

*ANA: [<] <pero t ú no sientes> ?

*ELE: y el pobre me dec ía [/] me dec ía / pero [/] pero /
qué pasa ? o sea no te importo nada ? <y yo>

*ANA: [<] <claro> //

*ELE: / le dec ía / o sea / mira / bueno / para empezar /
si no me
importara / no estaba ahora mismo aquí // eso / para empezar //
porque he estado contigo + no /<perdona> /

*ANT: [<] <hhh> //

%act: (1) laugh

*ELE: / pero es verdad //

*ANA: <t ía / es la primera> +

*ELE: [<] <o sea / yo la primera> + oye / yo la [///] perdonad un
momento // yo la primera vez que me enroll é con Fran / fue en
noviembre //

6.3. Corpus anotados a nivel semántico-pragmático

En lo que respecta al nivel semántico-pragmático de la lengua, solamente podemos destacar en el ámbito del español dos corpus que tienen anotación semántica. En primer lugar, el corpus Cast3lb (NAVARRO y otros, 2003), etiquetado con Named Entities, papeles semánticos (que se corresponden con su respectiva función sintáctica) y relaciones anafóricas; y, en segundo lugar, nuestro corpus, C-ORAL-ROM, anotado con una semántica eventiva (ALCÁNTARA, 2007) y también con información sobre expresiones referenciales (ALCÁNTARA y BERTOMEU, 2005). En el corpus Cast3lb, de la Universidad de Barcelona, las unidades que se encuentran en el nivel supraoracional se reconocen pero no se clasifican semánticamente. Los corpus con marcadores discursivos localizados proceden de la ingeniería lingüística, y se han etiquetado con miras al resumen automático o terminológicos (ALONSO y otros, 2002b), (ALONSO y otros, 2002a), (ALONSO y otros, 2003), (BACH, 2005) y (PRADA y MONCECCHI, 2003).

A pesar de que en lengua española los corpus etiquetados en estos niveles sean bastante escasos, no podemos dejar de señalar en el plano internacional las iniciativas casi ya de forma generalizada que se están realizando en todo el mundo en el etiquetado de aspectos que corresponderían a la dimensión pragmática. Dentro de la etiqueta de anotación pragmática encontramos modelos de anotación que comprenden estos temas: estructura retórica de los textos (PAGGIO, 2006) (CARLSON y otros, 2001), rasgos de oralidad (RODRÍGUEZ y otros, 2001) (ANDERSON y otros, 2004), argumentación (TEUFEL y MOENS, 1999), acuerdo, modalidad (NURMI, 2007), opinión (STOYANOV y CARDIE, 2008), estructura dialógica, actos de habla y prosodia, y por último emociones y sentimientos (ESULI y URCIUOLI, 2008), (LASKOWSKI y BURGER, 2006), (LASKOWSKI y BURGER, 2005), (GENEREUX y SANTINI, 2007) (MARTIN y otros, 2006), (DEVILLER y otros, 2006). Sin descartar los artículos que entienden por anotación pragmática la anotación de anáforas (NAVARRO y otros, 2003) (POESIO y ARSTEIN, 2008).

Hay un gran interés por parte de la comunidad científica tanto por anotar lo que se ha denominado marcas de subjetividad en las lenguas como por la estructura de los textos, como de hecho se pudo comprobar en los congresos LREC (Language Resources and Evaluation Conference) del 2008 y Corpus Linguistics de 2007, donde un buen número de comunicaciones estaban dominadas por estos temas. Sin embargo, de nuevo, los esquemas de anotación heredan esta falta de unificación teórica, por un lado, y por otro, de una definición de anotación pragmática unitaria, ya que encontramos anotaciones de fenómenos como la opinión, sin que se especifique claramente en qué nivel de lengua nos encontramos y si estamos hablando de un fenómeno lingüístico o una función textual.

Además, una crítica generalizada que puede hacerse a estos esquemas de anotación es que se centran demasiado en la anotación manual, no tienen un modelo teórico lingüístico detrás y no suelen asociar en sus trabajos fenómeno pragmático etiquetado con estructura lingüística, reflexión necesaria previa a la automatización.

Hace ya tiempo que la Lingüística de Corpus ha destacado las ventajas de la anotación de un mismo texto en varios niveles por encima de las diferentes versiones de un mismo texto en diferentes niveles. Esto ha originado una serie de estudios dedicados a mostrar este tipo de anotación que se ha denominado **anotación multinivel** y que fundamentalmente podemos encontrar en los corpus destinados la comunicación hombre-máquina (SORIA y otros, 2000) y (TAULÉ y otros, 2008) (Kirk, 2005), (WEISSER, 2004).

Dentro de las herramientas de anotación, es importante destacar la que ha diseñado M. Odonell (O'DONELL, 1997) para anotar estructura retórica de los textos, de libre descarga, lo cual es digno de destacar para los que defienden que la investigación pública debe ser como su propio adjetivo indica, pública. Por otra parte, también debemos destacar la herramienta Spaacy, diseñada para anotar un texto en diferentes niveles de anotación (WEISSER, 2004).

Por último, y para terminar el repaso en materia de anotación pragmática, es de rigor recordar que estas propuestas parten de un planteamiento equivocado desde un punto de vista epistemológico. Van a la caza de la subjetividad de determinados enunciados (WIEBE y otros, 2004) (WILSON, 2008), sin plantearse que dicha subjetividad es de raíz cognitiva, que impregna todos los enunciados, independientemente de si la opinión se codifica verbalmente o no. La separación que los periodistas suelen hacer entre información y opinión es falsa. La propia conceptualización de los fenómenos de la realidad ya está determinada por una visión subjetiva de los mismos. Como ejemplo de la no diferenciación podemos citar la posición de los locutores de radio españoles ante la selección de España, a la hora de definir qué es penalty y qué no; o el análisis que G. Lakoff hace del lenguaje político en su obra *No pienses en un elefante*, y en concreto del término *alivio de impuestos* reproducido por los periodistas -que supuestamente informan-, que hacen concebir los impuestos desde su raíz como una carga. Así pues, llamamos la atención desde este trabajo sobre este fenómeno de la anotación de la subjetividad, y donde se parte del falso supuesto de que los enunciados en los que no hay una palabra que codifique emociones o una marca de modalidad tienen más objetividad que los otros.

Aunque sea evidente por lo expuesto hasta aquí, cabe recalcar para terminar esta parte que no existe un sistema de anotación pragmática estándar que comprenda una representación de todas sus teorías y de los fenómenos más representativos. Parece pues que la creación de un estándar sobre información pragmática de corpus parece todavía un de-

safío bastante lejano. En 2010 saldrá a la luz el Hanbook de Corpus Linguistics, con un capítulo dedicado a anotación pragmática, veremos cuáles son sus propuestas en este sentido.

Este modelo de anotación, tal y como se señaló en la introducción, es una iniciativa que puede inspirar la creación del estándar, no tanto en materia de formalización sino en lo que se refiere al tipo de información pragmática que se debe considerar (GONZÁLEZ-LEDESMA, 2007a).

6.3.1. Pragmatext en C-ORAL-ROM en formato XML

Este apartado está dedicado a explicar el formato de la transcripción de C-ORAL-ROM etiquetado en XML con Pragmatext. En el capítulo cinco ya hemos adelantado el modo en que aparecen etiquetadas las unidades de información pragmática en XML. Aquí damos una explicación global al texto de la transcripción.

Como ya sabemos, la información pragmática se ha colocado a nivel de atributos. De lo contrario, esto es, si hubiéramos decidido declarar los fenómenos pragmáticos a nivel estructural, hubiéramos tenido que predecir su coaparición, y esto hubiera complicado bastante el esquema. Por ejemplo, ¿qué hubiéramos puesto antes? ¿un acto de habla o una operación argumentativa? Además, esta información a veces se cruza entre sí: por ejemplo, una unidad fraseológica puede ser un marcador discursivo pero no todas las unidades fraseológicas son marcadores discursivos, ni todos los marcadores discursivos son unidades fraseológicas. Es por eso que hemos preferido dejar la parte de los elementos para la estructura del texto y la de los atributos, como decimos, para reflejar la información pragmática en rasgos sin jerarquizar, todos al mismo nivel. Además, en el caso de que en el futuro se decida reflejar otro tipo de información, este modelo nos permite añadir libremente más fenómenos de naturaleza pragmática sin tener que hacer grandes cambios en la DTD.

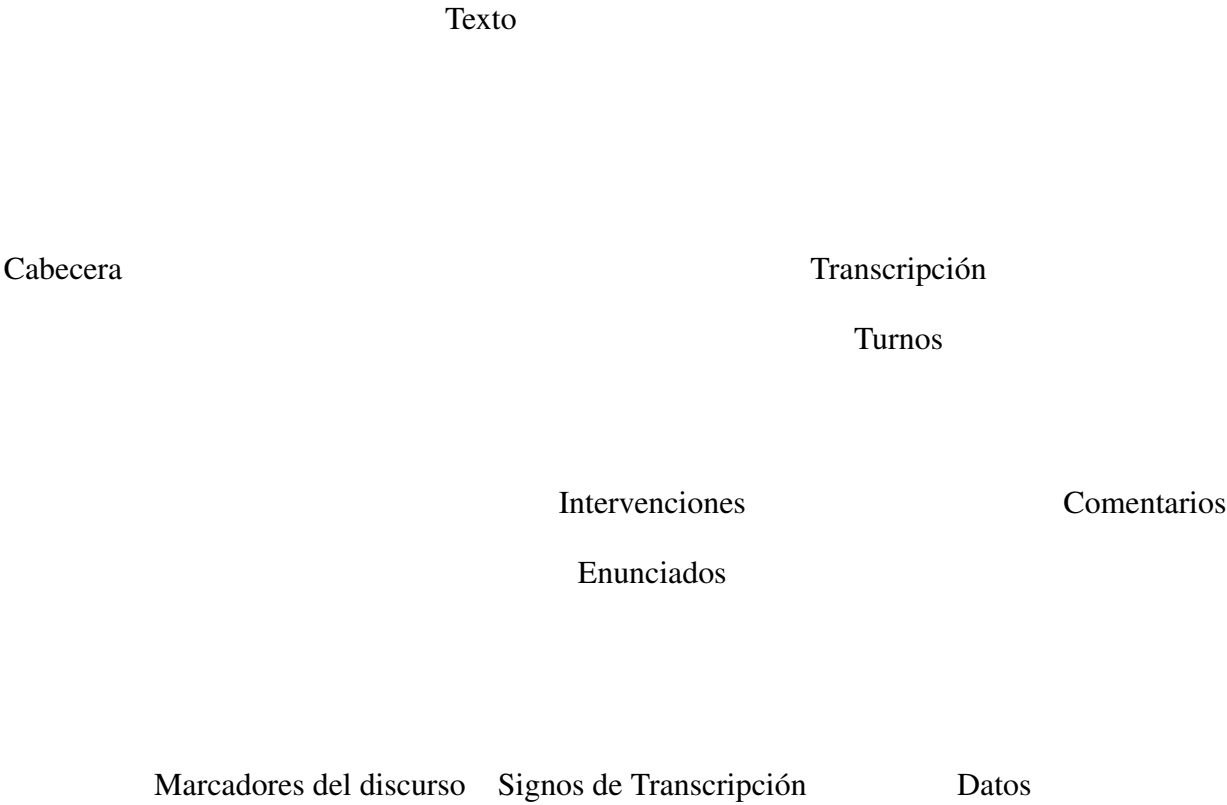
A nivel de elementos, nuestra formalización a XML respeta el formato CORALROM XML ¹. Usando las mismas etiquetas en XML que codificaron los textos vírgenes de la transcripción, hemos diferenciado cabecera y texto, y dentro de este, las intervenciones de los participantes junto con los comentarios o anotaciones. Cada intervención está compuesta por enunciados que tienen los atributos de ser interrogativos, aseverativos, interrumpidos, suspendidos o incompletos ². Cuando haya un marcador del discurso dentro de una inter-

¹Salvo que &eh y &ah dejan de ser etiquetados como apoyos vocálicos y pasan a ser unidades de planificación discursiva.

²Los cuales acabarán respectivamente en ?, //, +, ..., y ningún símbolo o [/], / en el caso de los finales de turno.

vención se le introducirá su correspondiente etiqueta.

A continuación presentamos un organigrama sobre la macroestructura de los textos, y seguidamente un ejemplo de transcripción etiquetada con el modelo de anotación Pragma-text.



```

<?xml version="1.0" encoding="UTF-8"?>

<!DOCTYPE Transcription SYSTEM "coralromMD.dtd">
<Transcription>
<Header>

<Title> problemas de novios </Title>
<File> efamcv10 </File>
<Participants>
<Speaker>
<ShortName> ELE </ShortName>
<FullName> Elena </FullName>
<Sex Type="woman" />
<Age Type="B" />
<Education Type="3" />
<Occupation> student </Occupation>
<Role> participant </Role>
<Origin> Madrid </Origin>
</Speaker>
<Speaker>
<ShortName> ANT </ShortName>
<FullName> Antonio </FullName>
<Sex Type="man" />
<Age Type="B" />
<Education Type="3" />
<Occupation> student </Occupation>
<Role> participant </Role>
<Origin> Madrid </Origin>
</Speaker>
<Speaker>
<ShortName> ANA </ShortName>
<FullName> Ana </FullName>
<Sex Type="woman" />
<Age Type="B" />
<Education Type="3" />
<Occupation> student </Occupation>

```

<Role> participant </Role>
 <Origin> Madrid </Origin>
 </Speaker>
 <Speaker>
 <ShortName> SEV </ShortName>
 <FullName> Ana </FullName>
 <Sex Type="woman" />
 <Age Type="B" />
 <Education Type="3" />
 <Occupation> student </Occupation>
 <Role> participant </Role>
 <Origin> Sevilla </Origin>
 </Speaker>
 </Participants>
 <Date> 03/04/2001 </Date>
 <Place> Madrid </Place>
 <Situation> chat between friends at university, hidden,
 researcher participant </Situation>
 <Topic> friends and lovers </Topic>
 <Source> C-ORAL-ROM </Source>
 <Class Type1="informal" Type2="family_private" Type3="conversation" />
 <Length> 6' 24'' </Length>
 <Words> 1568 </Words>
 <Acoustic_quality Type="A" />
 <Transcribers>
 <Transcriber> Manuel </Transcriber>
 </Transcribers>
 <Revisors>
 <Revisor> Guillermo </Revisor>
 <Revisor> Manuel and Jesús (prosody) </Revisor>
 </Revisors>
 <Comments> </Comments>
 </Header>

<Text>

<f h="ELE" file="efamcv10">
 <sf tipo="enunciation"> yo le entend ía perfectamente </sf>
 <sf tipo="suspension"> <DM grupo="2" tipo="2.2" FU="No"
 MET="No" SD="No" TD="No" DR="Coargumentation2" ED="No"
 MOD="No" EVI="No" SA="No" DEX="No">y</DM> lo
 <CAT grupo="0" tipo="0">que</CAT> no le quer ía decir era
 / decirle / <DM grupo="2" tipo="1.5" FU="No" MET="No"
 SD="No" TD="No" DR="No" ED="No" MOD="Interaction"
 EVI="No" SA="No" DEX="Social_Interlocutor">mira</DM>
 / Fran / <DM grupo="1" FU="Loc" MET="No" SD="No"
 TD="No" DR="Reformulation" ED="No" MOD="No" EVI="No"
 SA="No" DEX="No">o sea</DM> / yo no estoy enamorada de ti
 / <DM grupo="1" FU="No" MET="No" SD="No" TD="No"
 DR="Consequence" ED="No" MOD="No" EVI="No"
 SA="No" DEX="No">entonces</DM> no lo voy a </sf>
 </f>

<f h="ANA" file="efamcv10">
 <sf tipo="enunciation"> <DM grupo="2" tipo="1.5"
 FU="No" MET="S í" SD="Light" TD="Evidence" DR="No"
 ED="No" MOD="Truth_Booster" EVI="No" SA="No"
 DEX="No">claro</DM> </sf>
 </f>

<f h="ELE" file="efamcv10">
 <sf tipo="enunciation"> no le quer ía decir
 <DM grupo="2" tipo="3.4" FU="No" MET="No" SD="No"
 TD="No" DR="No" ED="No" MOD="Truth_Booster" EVI="No"
 SA="No" DEX="Social_Interlocutor">eso</DM> </sf>

<sf tipo="interruption"> <DM grupo="1" FU="No" MET="No"
 SD="No" TD="No" DR="Consequence" ED="No" MOD="No" EVI="No"
 SA="No" DEX="No">entonces</DM> no [/] no [/] no </sf>

<sf tipo="enunciation"> <DM grupo="1" FU="No" MET="No"
 SD="No" TD="No" DR="No" ED="Surprise" MOD="No" EVI="No"

SA="No" DEX="No">¡ah!</DM> / <DM grupo="1" FU="No" MET="No"
SD="No" TD="No" DR="Cause" ED="No" MOD="No" EVI="No" SA="No"
DEX="No">porque</DM> <DM grupo="1" FU="No" MET="No" SD="No"
TD="No" DR="Coargumentation2" ED="No" MOD="No" EVI="No"
SA="No" DEX="No">adem ás</DM> <DM grupo="2" tipo="3.1"
FU="Loc" MET="S í" SD="Existence" TD="Explanation"
DR="Explanation" ED="No" MOD="No" EVI="No" SA="Justification"
DEX="No">es que</DM> <DM grupo="1" FU="No" MET="No"
SD="No" TD="No" DR="Coargumentation2" ED="No" MOD="No"
EVI="No" SA="No" DEX="No">tampoco</DM>
sabía <CAT grupo="0" tipo="0">muy bien</CAT> lo <CAT grupo="0"
tipo="0">que</CAT>le quer ía decir </sf>
<sf tipo="interruption"> <DM grupo="1" FU="No" MET="No"
SD="No" TD="No" DR="Consequence" ED="No" MOD="No"
EVI="No" SA="No" DEX="No">entonces</DM> / a m í me pa </sf>
<sf tipo="continuation"> <DM grupo="1" FU="Col" MET="No"
SD="No" TD="No" DR="No" ED="No" MOD="Interaction" EVI="No" SA="No"
DEX="Social_Interlocutor">qu é pasa ?</DM> </sf>
</f>

<f h="ANT" file="efamcv10">
<sf tipo="enunciation"> hhh / <DM grupo="2" tipo="1.5"
FU="No" MET="No" SD="No" TD="No"
DR="Indefinite_Transitional_End" ED="No" MOD="Formulation_Hedge" EVI="No"
SA="No" DEX="No">nada</DM> </sf>
<Notes Type="act"> (1) laugh </Notes>
</f>

<f h="ELE" file="efamcv10">
<sf tipo="suspension"> <DM grupo="1" FU="No"
MET="No" SD="No" TD="No" DR="No" ED="Surprise"
MOD="No" EVI="No" SA="No" DEX="No">¡ah!</DM> /
<DM grupo="1" FU="No" MET="No" SD="No" TD="No"
DR="Consequence" ED="No" MOD="No" EVI="No"
SA="No" DEX="No">entonces</DM>
yo lo <CAT grupo="0" tipo="0">que</CAT>

hhh quer ía </sf> <Notes Type="act"> (6) laugh </Notes>
</f>

<f h="ANA" file="efamcv10">
<sf tipo="enunciation"> hhh </sf>
<Notes Type="act"> (1) laugh </Notes>
</f>
</Text>
</Transcription>

Aquí termina el capítulo seis. A continuación, en el siguiente capítulo profundizaremos en el reconocimiento, desambiguación y etiquetado automático de las unidades de información pragmática.

Capítulo 7

Estrategias de reconocimiento y clasificación semiautomática de los marcadores del discurso del corpus C-ORAL-ROM: el anotador DiMaSTagger y su evaluación

7.1. Introducción

En este capítulo, vamos a tratar todas las cuestiones relativas a la gestión computacional del corpus a la hora de reconocer estas palabras e introducir las etiquetas en XML que hemos diseñado para los marcadores en el corpus.

El tratamiento automático de los marcadores del discurso es una tarea de reciente en la investigación de Lingüística Computacional. Ya J. Teston y J. Veronnis (TESTON y VERÓNIS, 2004) han señalado que dicha categoría no se tiene en cuenta en los etiquetadores categoriales, con lo que debemos cuestionar los datos de cobertura y precisión de los mismos, sobre todo en lo que a la aplicación de dichos etiquetadores a corpus orales se refiere.

Las tareas en las que se ha centrado dicha investigación ha consistido en la identificación automática de marcadores del discurso en grandes corpus (MARCU, 1998) (POPESCU-BELIS y ZUFFEREY, 2006); su desambiguación y clasificación automática (FICHER y BRANDT-POOK, 1998) (PRADA y MONCECCHI, 2003) (HUTCHINSON, 2005) y, por último, la programación de etiquetadores automáticos de marcadores del discurso (T'SOU y otros, 2000).

La presencia de marcas de estructura superficial en los textos, como por ejemplo, la división en párrafos o los signos de puntuación, así como la coaparición de otros marcadores

del discurso, juega un papel clave en la realización de estas tareas, como ya señalamos en el capítulo tres. En cuanto a los etiquetadores de marcadores del discurso solamente hemos encontrado en la red un etiquetador para el chino. Como decimos, esta preocupación de la Lingüística Computacional por los marcadores del discurso y su tratamiento automático es relativamente reciente como demuestran la fecha de las tesis de Hutchinson, 2005, y los estudios de A. Popescu y S. Zufferey, del 2006. Recomendamos especialmente estos dos últimos trabajos para un estado de la cuestión en métodos de Lingüística Computacional en la identificación y clasificación automática de estas partículas en los textos.

Por lo que se refiere a la exposición de los conocimientos que se presentan a continuación, hemos estructurado este capítulo en cuatro partes:

1. Fase de preanotación.
2. Fase de anotación.
3. Fase de evaluación de la anotación.

7.2. Fase de preanotación

7.2.1. Elaboración del lexicon de marcadores del discurso

Nuestro lexicon de marcadores discursivos se ha elaborado con las siguientes fuentes:

1. Las obras de J. Portolés, L. Cortés, M^a.A. M. Zorraquino y C. Fuentes (PORTOLÉS, 1998c), (CORTÉS y CAMACHO, 2005), (CORTÉS, 1995a), (CORTÉS, 1995b), (MARTÍN y PORTOLÉS, 1999), (FUENTES, 1997a), (FUENTES, 1998).
2. El lexicon de interjecciones del DRAE (Real Academia~Española, 2000).
3. El lexicon que durante la etapa del etiquetado automático morfosintáctica del corpus y su validación manual había originado discusiones dentro del grupo de investigación sobre su categoría gramatical y, por tanto, sobre su significado. Ejemplos claros son *es que*, *madre mía de mi vida y de mi corazón* y *o sea*, los cuales no admiten *Part of Speech* sin estar faltando al funcionamiento de la gramática, ya que estas palabras son multiwords y ya no tienen significado composicional.

7.2.2. Reconocimiento de los candidatos a MD en el corpus

Tenemos que tener en cuenta que esta lista que hemos elaborado en realidad no es una lista de marcadores del discurso sino que es una lista de palabras que son candidatas

posibles a que en el corpus sean realmente marcador discursivo. Así por ejemplo, algunas palabras de la lista no estarán, otras estarán pero posiblemente con otro significado, como puede ser el caso de *bien*, y otras estarán pero a veces serán esta categoría y en otras ocasiones cumplirán otras funciones gramaticales, como es el caso de *bueno*.

Nuestro corpus tiene 300.000 palabras, y el objetivo de la tesis es etiquetar en ella los marcadores del discurso, con lo cual, no podemos arriesgarnos a definir a priori cuáles se comportarán como marcadores del discurso y cuáles no sin conocer previamente el posible nivel de ambigüedad en el corpus. En principio, podemos albergar la certeza de que en un corpus oral, palabras como *o sea* siempre sean marcador y no la suma de *o* más *sea*, sin embargo, esta certeza disminuye con otras palabras como *hombre o bien*. Con lo cual, para asegurarnos de que dividimos correctamente a los candidatos a la lista en los siguientes tipos:

1. No ambiguos¹
2. Ambiguos categoriales²
3. Ambiguos semánticos³
4. Ambiguos categorial y semánticamente⁴

hemos diseñado el programa **saltomortal.pl**. Cada vez que el programa encuentre una palabra de la lista de candidatos a MD en el corpus, imprimirá toda la línea en la que esta palabra aparece a un documento que tenga el mismo nombre que el marcador. El resultado es que el lingüista, una vez que pasa este programa, se encuentra con una carpeta que contiene archivos que se titulan como el marcador cuyo nivel de ambigüedad va a analizar. Dentro de cada archivo, se encuentran imprimidas las líneas en las que dicho marcador aparece en el corpus. De esta manera, cuando el lingüista lea los datos podrá analizar y

¹492 marcadores conforman la lista de no ambiguos, después de un gran esfuerzo por reducir marcadores que la tradición considera multifuncionales como por ejemplo *bueno* a una sola etiqueta (en este caso, atenuante) que explique todos los casos. Lo mismo se podría decir para marcadores como *pues* o *entonces*. En el interfaz web tratamos esta polémica que por problemas de espacio no hemos podido reflejar aquí.

²Conformado por 121 marcadores. Advertimos, estos marcadores son ambiguos **en nuestro corpus**. La ambigüedad puede cambiar según los datos. Algunos lingüistas computacionales deben comprender que la ambigüedad no puede tratarse en términos absolutos.

³A continuación, enumeramos los marcadores que son ambiguos semánticos: *coño*, *joder!*, *igualmente*, *luego*.

⁴Grupo conformado por *bien*, *joder*, *digo*, *venga*, *como*, *justo*, *igual* y *hombre*. Un lingüista diría que es una clasificación en términos de ambigüedad muy reduccionista. En cambio, es un gran paso desde un punto de vista computacional la distinción de los marcadores en estos grupos, y el diseño de estrategias computacionales diferentes en función del tipo de ambigüedad. Las listas de marcadores se pueden consultar en el CD.

evaluar el grado de dificultad que ese marcador va a tener a la hora de etiquetarlo y lo clasificará, como decimos, en uno de estos cuatro grupos. Además, también le hemos dicho al programa que cuando no encuentre un MD que lo imprima en otro documento diferente, llamado *losqueno.txt*, para que en el futuro dichos marcadores se tengan en cuenta a la hora de buscar en otros corpus diferentes a C-ORAL-ROM.

Al final ya tenemos completas las cuatro listas; con cada una de ellas vamos a seguir una estrategia diferente a la hora de localizar y etiquetar estas palabras en el corpus ⁵.

Con estas listas de marcadores hacemos lo siguiente con cada uno de los marcadores que las contienen:

1. Se consulta la bibliografía sobre el significado de esta partícula.
2. Se elaboran en documento Excell las etiquetas según el modelo de anotación, y luego, mediante en el programa que hemos diseñado (de csv a xml.pl) estos textos se pasan a formato XML. Ahora ya tenemos cuatro documentos en XML con las etiquetas correspondientes, y que podemos consultar en los anexos. El resultado de esta parte del trabajo está reflejada en el capítulo cinco.

Al final de esta etapa, ya sabemos cuáles son los marcadores de C-ORAL-ROM y ya tenemos una lista de etiquetas que nos hablan de los fenómenos que intervienen en su significado.

Como se podrá observar, esta etapa que precede al anotador, donde se han reconocido y clasificado semántico-pragmáticamente con nuestro modelo de anotación los marcadores, se caracteriza fundamentalmente por ser una tarea ad-hoc, y dista mucho del tratamiento que en Lingüística Computacional recibe esta pre-etapa, donde como ya hemos dicho, tanto D. Marcu como L. Alonso utilizan la prosodia para extraer automáticamente marcadores del discurso. Es justo señalar que para operar como lo han hecho estos investigadores necesitamos un gran volumen de palabras. D. Marcu trabaja con el Brown Corpus, que tiene aproximadamente un millón de palabras. Además el objetivo de estas investigaciones es mejorar las técnicas automáticas de identificación de estas partículas en grandes volúmenes de información, mientras que el objetivo de este trabajo es profundizar en el conocimiento de dichas partículas a partir de su estudio en un corpus oral, por lo que nos debemos asegurar de que están todos los marcadores del discurso que son, no podemos

⁵Pueden consultarse las listas y el programa en sus correspondientes anexos del CD.

decir que hemos localizado el 75% de los mismos, ya que luego vamos a usar el corpus para estudios de carácter cuantitativo y para la enseñanza de español a alumnos extranjeros.

7.3. Fase de anotación

El anotador automático DiMaSTagger está dividido en dos partes: un reconocedor y un etiquetador. El reconocedor busca, identifica candidatos a marcador discursivo en los textos; mientras que el etiquetador introduce la etiqueta en Xml correspondiente dentro del texto, una vez que el reconocedor ya ha hecho su trabajo.

Como ya hemos señalado anteriormente, hay cuatro tipos de marcadores del discurso. Hemos clasificado los cuatro tipos en función de su nivel de dificultad, caracterizándolos según este parámetro, fácil/difícil. La tabla arriba presentada nos indicará más didácticamente nuestro planteamiento.

Tipo de marcador	Reconocimiento	Etiquetado
No ambiguos	Fácil	Fácil
Ambiguos Catoriales	Difícil	Fácil
Ambiguos Semánticos	Fácil	Difícil
Ambiguos Semánticos-Catoriales	Difícil	Difícil

Tabla 7.1: Clasificación de las etapas del etiquetado por dificultad

Explicemos la tabla más detalladamente:

1. Los marcadores no ambiguos son fáciles tanto de reconocer como de etiquetar, ya que, como su propio nombre indica, tienen una sola categoría y una sola etiqueta, con lo que un simple programa en Perl basado en expresiones regulares soluciona el problema.
2. Los ambiguos categoriales son difíciles de reconocer, porque pueden ser o bien marcador del discurso o bien otra categoría, con lo que hay que hacer reglas de desambiguación contextual que diferencien uno de otro, para que después etiquete solo los que son marcadores discursivos. Sin embargo, por contrapartida, la etapa de etiquetado en esta clase de marcadores será una etapa fácil, porque una vez desambiguados sólo les correspondería una etiqueta, y el programa que introduce etiquetas en palabras no ambiguas es fácil de hacer.
3. Los ambiguos semánticos son fáciles de reconocer, porque todas las ocurrencias en el texto serían marcador del discurso, por tanto, solo tienen una categoría; pero, sin embargo, son difíciles de etiquetar porque tienen más de una etiqueta pragmática en XML.
4. Y por último, los semánticos-categoriales son tanto difíciles de reconocer como de etiquetar por las razones que ya hemos explicado, esto es, tienen más de una categoría y más de una etiqueta.

Esta es la razón por la que DiMaSTagger es un programa estructurado en cuatro módulos o subrutinas; dos módulos para el reconocedor, uno para los fáciles de reconocer (Módulo Fácil) y otro para los difíciles de reconocer (Módulo Difícil); y dos módulos para el etiquetador, el primero para los que son fáciles de etiquetar (Módulo Fácil), y el segundo para los que son difíciles de etiquetar (Módulo Difícil). Aun corriendo el riesgo de ser redundantes, presentamos arriba una tabla para asegurarnos de que se entiende el planteamiento del programa.

Tipo de marcador	Reconocimiento	Etiquetado
No ambiguos	Módulo Fácil	Módulo Fácil
Ambiguos Catoriales	Módulo Difícil	Módulo Fácil
Ambiguos Semánticos	Módulo Fácil	Módulo Difícil
Ambiguos Semánticos-Catoriales	Módulo Difícil	Módulo Difícil

Tabla 7.2: Separación del programa DiMaSTagger en módulos de dificultad según el tipo el tipo de marcador

La abstracción del problema y su separación en partes, tiene, de cara a la lógica de programación, consecuencias muy positivas, ya que la estrategia a seguir para etiquetar cada grupo de marcador es una combinación entre los cuatro módulos, optimizando como consecuencia todo lo posible el código.

Pasemos ahora a explicar los módulos. Los módulos fáciles, tanto los de reconocimiento y como los de etiquetación, no vamos a explicarlos, ya que, son fáciles de programar, solo consiste en localizar una expresión regular en el texto (que se correspondería con la cadena de caracteres del MD) y añadirle una etiqueta. Centrémonos por tanto en la explicación de los módulos difíciles.

7.3.0.1. El módulo de reconocimiento y de desambiguación contextual

Este módulo ha sido diseñado para los dos grupos de marcadores que contienen ambigüedad categorial. Si queremos automatizar el proceso de anotación lo máximo posible, debemos establecer algún tipo de criterio que, *a priori*, nos ayude a decidir por una opción (marcador) o por otra (otra categoría). Llegados a este punto, hemos decidido servirnos de las marcas prosódicas en el texto como posibles indicadores de función, habida cuenta de la importancia de las marcas de entonación en la semántica de los marcadores del discurso, en la coherencia textual y en la interpretación del significado, mencionada en más de una ocasión tanto por la Pragmática Teórica ⁶ como por la Pragmática Computacional.

Como ya explicamos en el capítulo seis, estas son las marcas prosódicas que codifican pausas y tipos de entonación en C-ORAL-ROM: /, //, ?, ..., +. Nuestra hipótesis de partida es que un candidato a MD tendrá un grado de probabilidad alto de serlo si se encuentra rodeado por un lado, por otro o por los dos lados, tanto de una marca prosódica como de un marcador del discurso ⁷. Ahora bien, implementar este enunciado-hipótesis en el programa de ordenador no es fácil. Fijémonos en que estamos diciendo que será marcador del discurso si a la izquierda aparece o bien una marca prosódica o bien un marcador o bien los dos, y a la derecha, o bien un marcador o bien una marca prosódica o bien los dos. Y también será marcador discursivo a priori un candidato que a la izquierda tenga al menos alguna de estas dos cosas, aunque a la derecha no tenga nada, y viceversa. La racionalización previa y necesaria al diseño de todo programa, esta que debe llevarse a cabo durante la concepción del algoritmo, no es baladí, ya que estamos manejando un grado elevado de

⁶La prosodia es uno de los mecanismos fundamentales que se utilizan oralmente para conseguir la producción de textos coherentes (TUSÓN, 1997) (WICHMANN y BLAKEMORE, 2006).

⁷La estrategia de reconocimiento de marcadores a través de la co-ocurrencia de otros marcadores ha sido utilizada como ya hemos mencionado por Hutchinson (HUTCHINSON, 2005) para la clasificación automática de marcadores.

posibilidades y los lenguajes de programación funcionan mejor con contextos fijos que con contextos cambiantes y/o opcionales como es nuestro caso. Así pues, para conseguir una buena abstracción del problema y asegurarme de llevar a cabo una implementación correcta del mismo, recurrí a la experiencia acumulada en diseño de programas para anotación automática de corpus y diccionarios del investigador de F. Sánchez, actualmente jefe de departamento de Lingüística Computacional e Informática de la RAE. A él debemos tanto el planteamiento teórico del problema como la parte del código en Perl correspondiente a dicho módulo del etiquetador automático, y por tanto, es de rigor citar aquí la autoría de su contribución, del todo altruista, a este trabajo.

El módulo de reconocimiento estará compuesto por las siguientes reglas contextuales, cada una de ellas presenta una posibilidad de combinación entre marcas prosódicas, marcadores del discurso y contexto a la izquierda y a la derecha. Cada regla respeta un principio sagrado para los programadores expertos que trabajan a menudo con expresiones regulares y que consiste en distinguir en cada regla o expresión regular tres partes claramente diferenciadas y que en la implementación se corresponderán con tres variables diferentes:

CTXTO IZQDA FOCO CTXTO DRCHA

A continuación, presentamos las reglas que se implementarán en el módulo de reconocimiento, y que afectará a los marcadores de tipo dos y de tipo cuatro.

Tipo	CI	FOCO	CD
1.1	MD TM	FOCO	TM MD
2.5	MD TM	FOCO	MD TM
2.6	TM MD	FOCO	TM MD
1.2	TM	FOCO	TM MD
1.3	MD	FOCO	TM MD
1.4	MD TM	FOCO	TM
1.5	TM	FOCO	TM
1.6	MD	FOCO	TM
1.7	MD TM	FOCO	MD
1.8	TM	FOCO	MD
1.9	MD	FOCO	MD
3.3	Ø	FOCO	MD TM
3.4	Ø	FOCO	TM MD
3.1	Ø	FOCO	MD
3.2	Ø	FOCO	TM
2.3	MD TM	FOCO	Ø
2.4	TM MD	FOCO	Ø
2.2	TM	FOCO	Ø
2.1	MD	FOCO	Ø
2.7	TM MD	FOCO	MD
2.8	TM MD	FOCO	TM
0	Ø	FOCO	Ø

Tabla 7.3: Tabla de especificación de los tipos contextuales, donde TM es transcription mark; y MD, discourse marker.

Cuando el etiquetador reconozca un marcador discursivo ambiguo categorialmente leerá su contexto y le aplicará un tipo de regla determinada. A continuación, sacará el texto por pantalla y el anotador humano comprobará si efectivamente la regla ha fallado o no. De esta manera, la anotación y la validación se harán conjuntamente en el espacio y el tiempo. Posteriormente, cuando hablemos del funcionamiento general del etiquetador, explicaremos más detenidamente este último aspecto.

7.3.0.2. Módulos de etiquetación

El módulo difícil de etiquetación afecta a aquellos grupos que tengan más de una etiqueta. En este caso, es imposible con la información de la que partimos saber cuándo una palabra como *hombre* (cuando es MD y no nombre) es un atenuante o bien es un codificador de la sorpresa. Es por eso que para ello debemos recurrir a la desambiguación humana. Así pues cada vez que el etiquetador se encuentre con un marcador del discurso con más de una etiqueta, el anotador automático ofrecerá las etiquetas asociadas al marcador, y el anotador humano seleccionará una de ellas, que será con la que definitivamente el programa anotará el marcador.

7.3.1. Funcionamiento a tiempo real de DiMaSTagger

Hemos hablado hasta aquí de un programa concebido en módulos que se combinan entre sí para dar cuenta de las especificidades de cada grupo de marcadores. Expliquemos ahora, también brevemente, cuál es la dinámica general del anotador.

El anotador anota texto por texto los marcadores de los cuatro grupos. De nuevo este planteamiento nos ha parecido el más óptimo, frente al de etiquetar todos los textos con el primer grupo, luego todos los textos de nuevo con el segundo grupo, validarlo, luego pasarle el programa que etiqueta los del tercer grupo y así sucesivamente. Con la estrategia *texto a texto* llevamos perfecta cuenta de lo que hemos etiquetado y de lo que nos queda por etiquetar, ya que el programa automáticamente selecciona un texto de la carpeta de corpus, y cuando el lingüista lo ha terminado de anotar, automáticamente lo pasa a la carpeta de anotados y selecciona el texto siguiente. Enumeremos con más detalle cada uno de los pasos de la estrategia *texto a texto*:

1. Se ejecuta el program perl main.pl
2. Se carga la lista total de marcadores ordenados alfabéticamente y en orden descendente, para que coja la forma *qué bien* antes que la forma *bien*, por ejemplo.

3. Elimina marcas de solapamiento y transforma los apoyos vocálicos *&eh* *&ah* en eh ah para ser reconocidos como marcadores.
4. Por línea de comandos también se advierte del nombre del archivo con el que se va a trabajar.
5. En este punto el programa despliega una ventana que va a ser el interfaz con el que vamos a trabajar, y que presentamos a continuación.

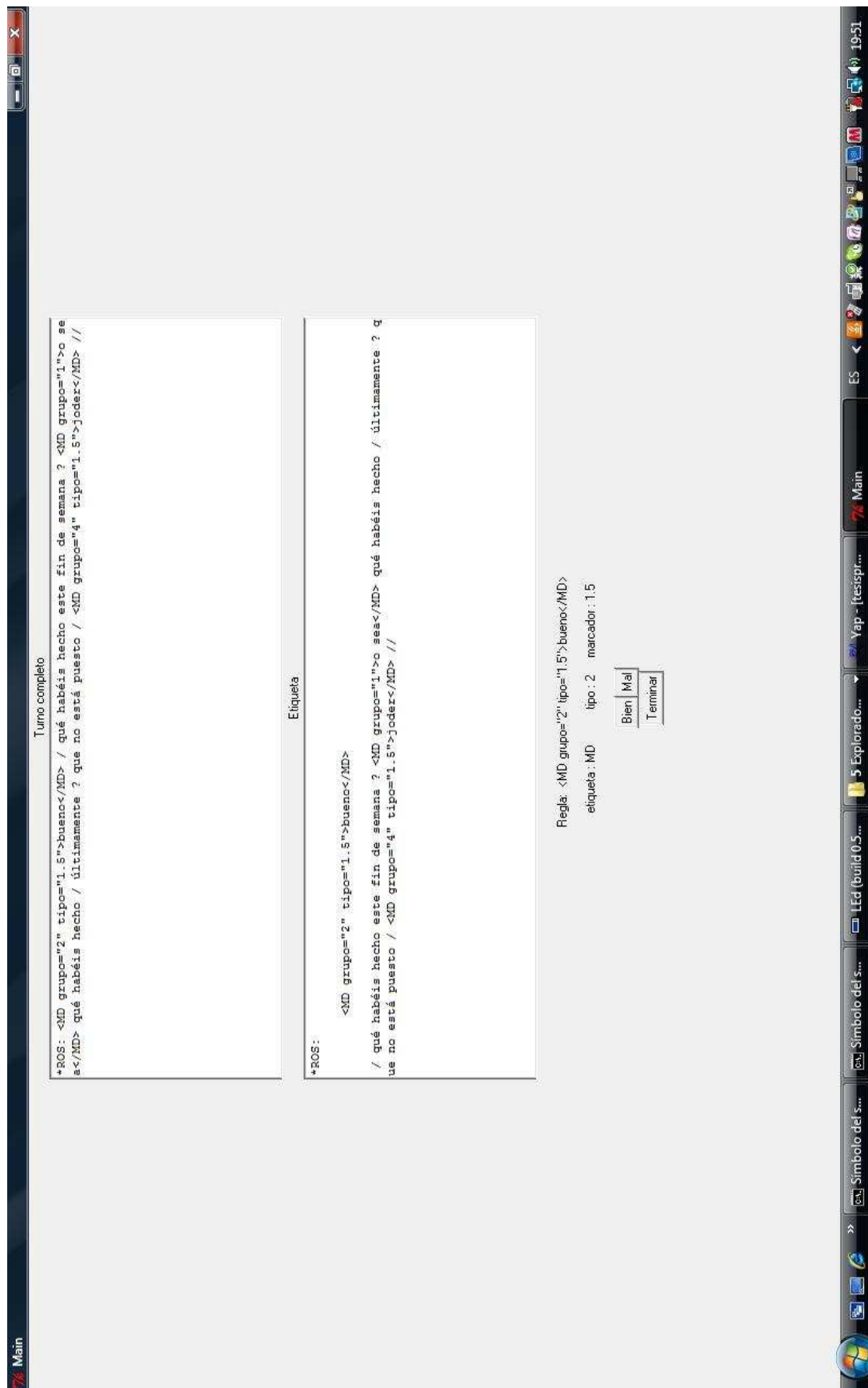


Figura 7.1: Ejemplo del anotador en la fase de desambiguación categorial

El programa reconoce el primer marcador. Si es ambiguo categorialmente sale por pantalla el marcador inserto en el turno marcado con la regla contextual que se ha aplicado y a continuación tres opciones: 1. Bien 2. Mal 3. Terminar.

6. El anotador humano marcará alguna de las tres opciones. Tanto si marca 1 como 2, estos datos se envían a un archivo externo, dónde se van contando los casos en los que el anotador automático ha acertado y los casos en los que ha fallado. Si marca mal se invertirá automáticamente la etiqueta y de CAT se pasará a MD y de MD a CAT; entendiendo por CAT otra categoría diferente a MD.
7. Una vez desambiguados todos los marcadores ambiguos, el programa pasa al módulo de desambiguación de etiquetas, y editará de nuevo una pantalla donde salga el marcador, el contexto y las etiquetas. El anotador humano pulsará 1, 2 o 3 dependiendo del número de etiquetas seleccionadas. A continuación el programa etiquetará con la etiqueta elegida. A continuación, presentamos una imagen del etiquetador en este punto del procesamiento:

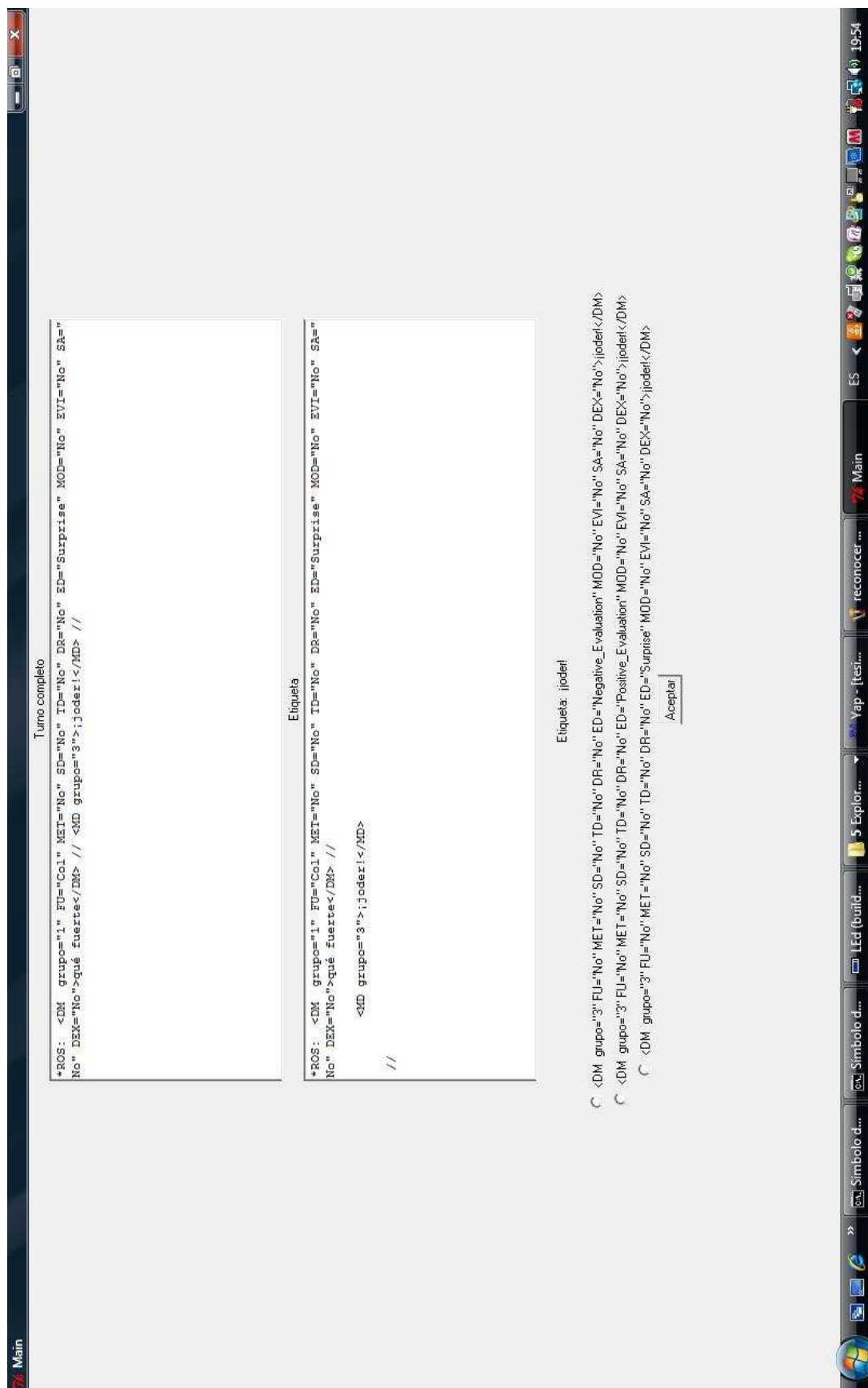


Figura 7.2: Ejemplo del anotador en la fase de desambiguación de etiquetas

8. Después de que los marcadores hayan sido reconocidos, validados y etiquetados, el resto del texto se codificará a XML ⁸ y se pasará automáticamente a la carpeta de etiquetados.
9. El anotador avisa de que comienza con el siguiente archivo.

Las fases por tanto son reconocer, corregir, etiquetar, y codificar a XML la transcripción.

7.4. Fase de evaluación de la anotación

Veamos a continuación las tablas en las que se reflejan las frecuencias de acierto y de error de las reglas contextuales y de los marcadores.

7.4.1. Resultados del grado de acierto de las reglas contextuales

⁸La fase de validación del corpus en XML se hace automáticamente, mediante un parser.

CI	FOCO	CD	TIPO	BIEN	MAL
TM	FOCO	TM	1.5	2030	114
TM	FOCO	Ø	2.2	962	501
TM	FOCO	TM MD	1.2	845	50
TM MD	FOCO	TM	2.8	433	25
MD TM	FOCO	TM	1.4	392	13
Ø	FOCO	TM	3.2	301	628
TM MD	FOCO	Ø	2.4	295	108
MD TM	FOCO	Ø	2.3	212	91
TM	FOCO	MD	1.8	190	19
MD TM	FOCO	TM MD	1.1	186	8
TM MD	FOCO	TM MD	2.6	154	7
Ø	FOCO	TM MD	3.4	154	278
MD	FOCO	TM	1.6	76	9
MD	FOCO	Ø	2.1	63	36
TM MD	FOCO	MD	2.7	46	7
MD	FOCO	TM MD	1.3	28	5
MD TM	FOCO	MD	1.7	26	4
MD TM	FOCO	MD TM	2.5	21	0
Ø	FOCO	MD	3.1	19	30
Ø	FOCO	MD TM	3.3	12	22
MD	FOCO	MD	1.9	4	2

Tabla 7.4: Tabla de especificación de los tipos contextuales

Como podemos observar en tabla, las reglas están ordenadas según el criterio de frecuencia de aparición. Sirva esta tabla y las siguientes para establecer la relevancia de las reglas sobre todo en lo que a corpus orales se refiere.

A continuación, presentamos tres tablas que presentan las reglas que funcionan, las que definitivamente tienen resultados negativos, y las que presentan contextos ambiguos en términos de frecuencia.

CI	FOCO	CD	TIPO	BIEN	MAL
TM	FOCO	TM	1.5	2030	114
TM	FOCO	TM MD	1.2	845	50
TM MD	FOCO	TM	2.8	433	25
MD TM	FOCO	TM	1.4	392	13
TM	FOCO	MD	1.8	190	19
MD TM	FOCO	TM MD	1.1	186	8
TM MD	FOCO	TM MD	2.6	154	7
MD	FOCO	TM	1.6	76	9
TM MD	FOCO	MD	2.7	46	7
MD	FOCO	TM MD	1.3	28	5
MD TM	FOCO	MD	1.7	26	4
MD TM	FOCO	MD TM	2.5	21	0

Tabla 7.5: Tabla de especificación de los tipos contextuales que presentan buenos resultados en la desambiguación

CI	FOCO	CD	TIPO	BIEN	MAL
Ø	FOCO	TM	3.2	301	628
Ø	FOCO	TM MD	3.4	154	278
Ø	FOCO	MD	3.1	19	30
Ø	FOCO	MD TM	3.3	12	22

Tabla 7.6: Tabla de reglas contextuales que han dado un resultado negativo

CI	FOCO	CD	TIPO	BIEN	MAL
TM	FOCO	Ø	2.2	962	501
TM MD	FOCO	Ø	2.4	295	108
MD	FOCO	Ø	2.1	63	36
MD	FOCO	MD	1.9	4	2

Tabla 7.7: Tabla de especificación de los tipos contextuales con resultado ambiguo

En el futuro estos datos servirán para mejorar la cobertura y precisión de un desambiguador, esta vez sí, automático de marcadores del discurso. Pero adelantemos ya, producto de la observación de las cifras de las tablas, la debilidad de las reglas con contextos vacíos.

7.4.2. Resultados del grado de acierto sobre CAT vs. MD

Veamos a continuación los datos de ocurrencias de formas lingüísticas que en el corpus se usan o bien como marcadores o bien como otra categoría. Estos datos también se tendrán en cuenta a la hora de hacer reglas de desambiguación particulares para cada marcador.

MD	BIEN	MAL
bueno	1242	50
es que	1056	74
claro	734	31
vamos	399	165
vale	218	22
mira	182	18
a ver	176	63
muy bien	143	122
y nada	127	6
hombre	121	51
venga	117	19
y eso	107	57
por eso	103	11
digamos	95	5
tal	92	16
usted	82	112
gracias	78	39
encima	78	17
ya	75	55
en fin	71	1
primero	69	51
pues eso	67	7
vamos a ver	64	16
y ya	62	38
y todo	52	22
que si	49	50
total	45	7
la verdad	39	18
de verdad	34	4
hija	32	10
exactamente	31	17
macho	31	1
vaya	33	25
mujer	24	34
seguro	23	4
lo siento	23	1
anda	21	1
eso	20	54
nada	20	27
quiero decir	20	6
ya que	20	4

Tabla 7.8: Tabla de frecuencia marcador vs. categoría

MD	BIEN	MAL
segundo	19	25
de acuerdo	19	22
por un lado	19	2
bien	15	11
maja	15	4
aparte	14	6
esto es	12	68
ya sabes	12	10
lo que pasa	12	8
fuerte	11	21
justo	9	20
concretamente	9	3
cuidado	8	10
estupendo	8	7
ya te digo	7	14
escucha	7	9
genial	7	9
de acuerdo con	7	2
adelante	5	28
con todo	5	19
anteriormente	5	12
ahora	4	6
muy fuerte	4	4
a saber	4	4
maravilloso	4	1
una mierda	3	7
evidente	3	7
de entrada	3	4
para nada	3	3
ojo	3	1
que digamos	3	1
por ello	3	1
por otro	3	2
igual	2	13
tercero	2	10
hay que ver	2	3
guay	2	3
en total	2	3
espectacular	2	5
la hostia	2	2
lo que se dice	2	2

Tabla 7.8: Tabla de frecuencia marcador vs. categoría

MD	BIEN	MAL
no me digas	2	1
majo	1	2
hostias	1	2
particularmente	1	2
es espectacular	1	1
muy breve	1	1
es muy fuerte	1	1
por decir algo	1	1

Tabla 7.8: Tabla de frecuencia marcador vs. categoría

MD	BIEN	MAL
es que	1056	74
bueno	1242	50
claro	734	31
vamos	399	165
vale	218	22
mira	182	18
a ver	176	63
y nada	127	6
hombre	121	51
venga	117	19
por eso	103	11
digamos	95	5
tal	92	16
encima	78	17
en fin	71	1
pues eso	67	7
vamos a ver	64	16
y todo	52	22
total	45	7
la verdad	39	18
de verdad	34	4
hija	32	10
macho	31	1
seguro	23	4
lo siento	23	1
joder	22	0
anda	21	1
quiero decir	20	6
ya que	20	4
por un lado	19	2
maja	15	4
aparte	14	6
concretamente	9	3
de acuerdo con	7	2
ojo	3	1

Tabla 7.9: Tabla de aciertos de marcador sobre categoría

MD	BIEN	MAL
mujer	24	34
eso	20	54
esto es	12	68
justo	9	20
ya te digo	7	14
escucha	7	9
adelante	5	28
con todo	5	19
anteriormente	5	12
una mierda	3	7
evidente	3	7
igual	2	13
tercero	2	10
espectacular	2	5

Tabla 7.10: Tabla de fallos de marcador sobre categoría

MD	BIEN	MAL
y eso	107	57
muy bien	143	122
usted	82	112
gracias	78	39
ya	75	55
primero	69	51
y ya	62	38
que si	49	50
vaya	33	25
exactamente	31	17
nada	20	27
segundo	19	25
de acuerdo	19	22
bien	15	11
ya sabes	12	10
lo que pasa	12	8
fuerte	11	21
cuidado	8	10
estupendo	8	7
genial	7	9
ahora	4	6
muy fuerte	4	4
a saber	4	4
maravilloso	4	1
de entrada	3	4
para nada	3	3
que digamos	3	1
por ello	3	1
por otro	3	2
hay que ver	2	3
guay	2	3
la hostia	2	2
lo que se dice	2	2
en total	2	3
majo	1	2
hostias	1	2
particularmente	1	2
es espectacular	1	1
muy breve	1	1
es muy fuerte	1	1
por decir algo	1	1

Tabla 7.11: Tabla de frecuencia ambigua de marcador sobre categoría

7.5. Conclusiones

En este punto del trabajo, hemos resuelto el núcleo duro de la investigación. Hemos diseñado un modelo de anotación que ya está reflejado en los textos del corpus, el cual en este momento de la investigación ya está codificado en formato XML. Ahora ya solo queda saber qué cosas se pueden hacer con el corpus. Damos paso por tanto a la penúltima parte de este ejercicio intelectual que va de lo teórico a lo aplicado, pasando por una etapa de sistematización descriptiva y formalización de la información. Para cada aplicación hemos diseñado también las herramientas computacionales necesarias para la extracción y presentación de los datos de acuerdo con los objetivos teóricos, de naturaleza lingüística y didáctica respectivamente, de cada capítulo. Hemos diseñado una interfaz Web para la recuperación y presentación de los datos que nos interesen del corpus en función de sus aplicaciones. Una de las secciones de esta interfaz está dedicado al siguiente capítulo, donde veremos cómo un contador de frecuencias de marcadores del discurso en función de géneros discursivos nos abrirá las puertas de una disciplina aún emergente como es la de la Pragmática Cuantitativa.

Parte IV

De la aplicación

Capítulo 8

Los marcadores del discurso en C-ORAL-ROM, entre la oralidad y la escritura: un análisis de corte cuantitativo

It had always seemed to me that the linguistic system is inherently probabilistic in the grammar. M. A. K. Halliday. (HALLIDAY, 2004)

8.1. Introducción

Parece que hay un acuerdo bastante general dentro de la comunidad científica que se dedica al estudio del significado de los marcadores discursivos a la hora de señalar que tanto la aparición de dichas palabras o *multiwords* como sus valores o usos están inextricablemente unidos a factores de carácter extralingüístico como por ejemplo el tipo de registro, la modalidad (oral o escrita) o la situación comunicativa (por ejemplo, una clase en un colegio), entre otros (MARTÍN y MONTOLÍO, 1998), (CASADO, 1998), (MONTOLÍO y UNAMUNO, 2000). M^a Martín Zorraquino (MARTÍN y MONTOLÍO, 1998) señala también que la intersección entre marcadores discursivos y sociolingüística, es “una de las asignaturas pendientes cuya laguna poco a poco se va sufragando”.

Al mismo tiempo, parece también haber un acuerdo general a estas alturas del conocimiento de las lenguas en torno a la idea de que la perspectiva cuantitativa es necesaria para la descripción gramatical de las lenguas en tanto que constata la frecuencia de los fenómenos lingüísticos en función de las diferentes variables contextuales; especialmente en lo que respecta a la lengua oral, ya que como sabemos, debido a razones tanto tecnológicas como ideológicas, hasta el día de hoy, sigue siendo bastante menos conocida que la variante escrita de las lenguas (CORTÉS, 1988).

En lo que respecta a los marcadores del discurso, existen muy pocos estudios de corte cuantitativo en relación a los de naturaleza cualitativa. Es bastante posible que tanto la no-formalización de la información de los corpus como el desconocimiento de las herramientas computacionales para etiquetar corpus y extraer frecuencias automáticamente haya obligado a los lingüistas a sentirse más seguros en el terreno de la cualidad que en el de la cantidad a la hora de enunciar sus conclusiones sobre el significado de estas partículas en contextos restringidos. Aparte, debemos señalar que aquellos que se han atrevido a estudiar dichas partículas desde esta perspectiva obvian como si de información relevante no se tratara todo el proceso de gestión y extracción de la información, además de no explicitar debidamente las características de los corpus en los que han basado su estudio.

Por todas estas razones, hemos decidido que la primera aplicación del corpus C-ORAL-ROM anotado con los marcadores del discurso esté orientada a los lingüistas. En ella, exponemos los datos de las frecuencias de los marcadores discursivos del corpus en función de algunas variables contextuales como el género discursivo o el número de hablantes. Sirva esta tesis hasta este capítulo como un ejemplo de la metodología que se pudiera seguir para aquellos grupos de investigación dedicados a la Lingüística descriptiva que decidan etiquetar los marcadores discursivos de sus respectivos corpus con técnicas computacionales y con fines lingüísticos.

8.2. Oralidad y escritura en los marcadores del discurso de C-Oral-Rom

Verba volant scripta manent

Una de las críticas que podríamos hacerle a aquellos que sí que poseen el conocimiento computacional pero no la formación lingüística es que, dado que ya tienen la herramienta, muchos estudios se disponen a obtener datos de frecuencia absolutamente irrelevantes o vacíos de significado para un investigador con formación lingüística y cuyo propósito es confirmar una visión sobre la lengua a partir de los datos de un corpus. Chomsky, ironizando sobre las posibles consecuencias de una lingüística estadística, afirma que:

[Los aspectos estadísticos] no tienen nada que ver con la gramática; por ejemplo, seguramente no es una cuestión propia de la gramática del inglés el hecho de que *Nueva York* sea más probable que *Nevada* en el contexto *Vengo de* —. En general, me parece que la importancia dada a las consideraciones

estadísticas ha sido sobrevalorada en las discusiones recientes sobre teoría lingüística. Chomsky, 1965 en (ALCÁNTARA, 2007)

Así pues, los datos tienen que ser extraídos en función de un objetivo. En este capítulo, nuestra meta será reflexionar sobre la oralidad y la escritura a través de los resultados de frecuencias de los marcadores de nuestro corpus. Si recordamos lo dicho en el capítulo seis, en las cabeceras hay un campo semántico para especificar el género comunicativo en relación al diseño previo del corpus. Recordemos que dicha información viene especificada en el apartado @Class, y que se rellena a partir de las tablas 8.1 y 8.3 que volvemos a incluir en este capítulo.

Informal	Familiar/Private	Monologue Dialogue Conversation
	Public	Monologue Dialogue Conversation

Tabla 8.1: Distribución de la parte informal de C-Oral-Rom

Formal		
Formal in Natural Context	Media	Telephone
Political Speech	News	Private Conversation
Political Debate	Sport	Phone Call Services (ManMachine interaction)
Preaching	Interviews	
Teaching	Meteo	
Professional Explanation	Scientific Press	
Conference	Reportage	
Business	Talk shows political debate	
Law (through media)	Talk shows thematic discussions	
	Talk shows culture	
	Talk shows science	

Tabla 8.3: Distribución de la parte formal de C-Oral-Rom

Veamos, pues, qué nos puede aportar la relación entre marcador y género comunicativo a la descripción de las diferencias entre la modalidad oral y la modalidad escrita desde un punto de vista cognitivo. Dos van a ser las fuentes de conocimiento en torno al tema de la oralidad de las que vamos a partir aquí. En primer lugar, no podemos dejar de citar la obra de W. Ong, *Tecnologías de la palabra* (ONG, 1993), en la que dicho autor argumenta que el conocimiento y uso del invento de la escritura produjo y produce una serie de transformaciones en las estructuras cognitivas, culturales y sociales del ser humano. Con el descubrimiento de la escritura el ser humano pudo materializar el conocimiento, pararlo en el devenir del tiempo, y someterlo a análisis y a crítica.

A esto debemos añadir los planteamientos del profesor Portolés en investigaciones como *Escritura y marcadores del discurso* (en (PORTOLÉS, 2005)), donde sostiene que el nacimiento de ciertos marcadores discursivos ha tenido lugar en el seno de las modalidades escritas; ya que la relación con lo ya dicho es efímera y ocasionalmente recuperable en lo oral a través de la paráfrasis¹; y en cambio, es sostenida y absolutamente recuperable cuando el discurso es escrito. Citemos sus palabras al respecto:

La comprensión del discurso como un fenómeno esencialmente pragmático y no semántico, como proponen las escuelas textuales, permite destacar el valor de la escritura como un hecho fundamental en el nacimiento y la consolidación de muchos de los marcadores del discurso. La escritura facilita una memoria suplementaria que posibilita, entre otras cosas, estructuras informativas más complejas que las habituales del discurso oral - tema/rema, foco y elementos con función de marco -. La capacidad del discurso escrito de conservar el discurso previo al enunciado que se está leyendo es la que explicaría la complejidad de las instrucciones semánticas de muchos marcadores discursivos. Sólo la planificación que permite la escritura justifica su existencia.

Asimismo, la extensión de la alfabetización a toda la sociedad podría dar cuenta del mayor número de marcadores discursivos del español actual en relación con etapas anteriores de la lengua. Nunca las innovaciones escritas, o leídas en público en el caso de los medios de comunicación, han tenido un número igual de receptores que poseen una mente configurada por el conocimiento de la escritura. Si comprendemos y los marcadores discursivos como memes, en la concepción de la cultura del zoólogo Richard Dawkins, esto es, como una

¹Salvo en los casos de memorización privilegiada de cadenas de palabras como se da en algunas mentes brillantes.

especie de virus culturales que buscan un organismo en el que replicarse, nunca ha habido una población tan favorable a dejarse infectar (PORTOLÉS, 2005).

Veamos qué marcadores del discurso se prefieren en la parte informal, en la formal y en la de medios y comunicación, y veamos también si hay algún tipo de diferencia con respecto a ellos. Pero antes explicaremos la parte de la aplicación Web Inter Alia dedicada a los estudios cuantitativos.

8.3. Inter Alia para lingüistas

Si el usuario quiere entrar al cuenta frecuencias en función del género discursivo de la aplicación debe pulsar la segunda opción de la página Web, después introducirá el marcador que desee analizar y a continuación aparecerán los datos sobre su frecuencia en el corpus. Ejemplos de estos tres pasos son las imágenes presentadas abajo:

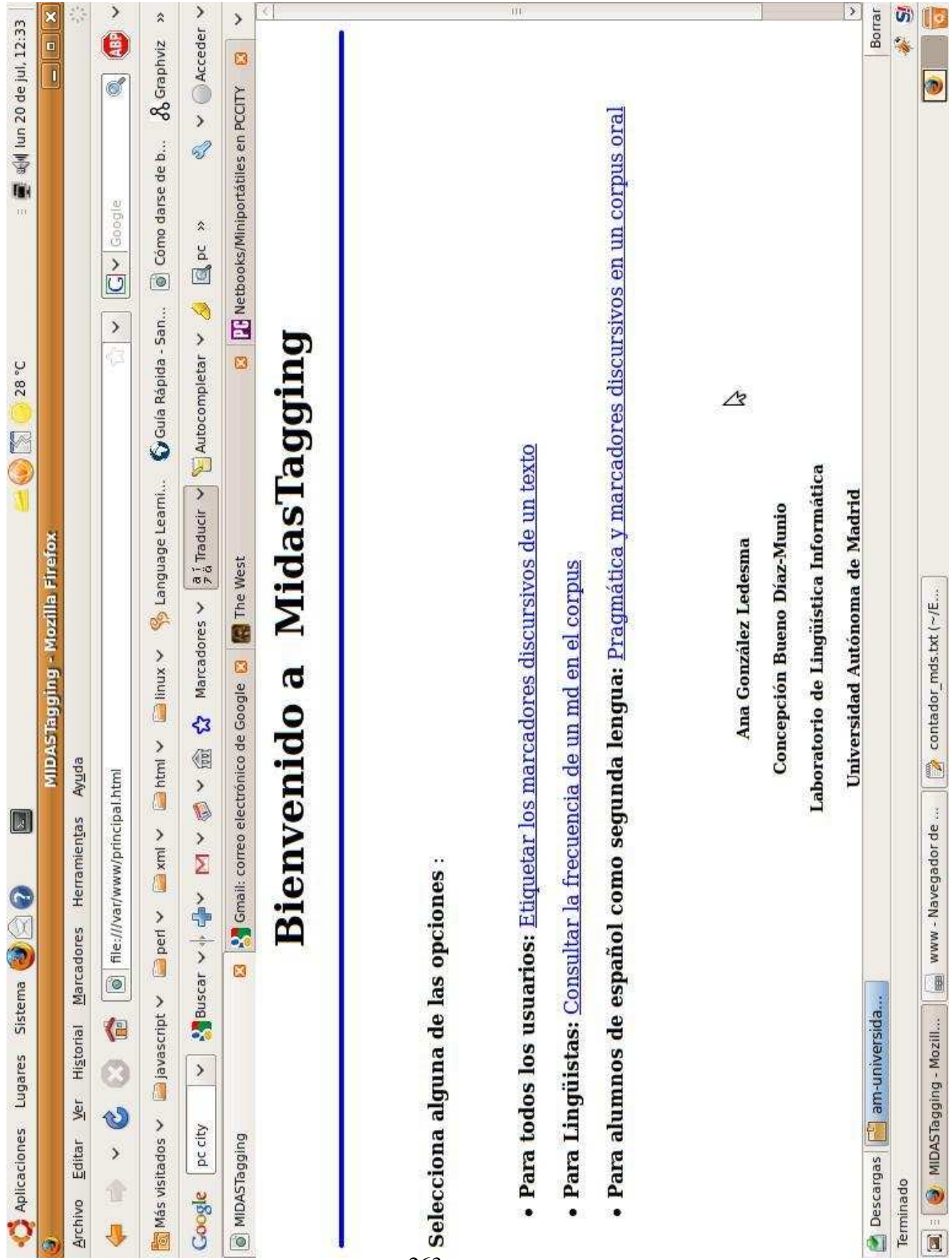
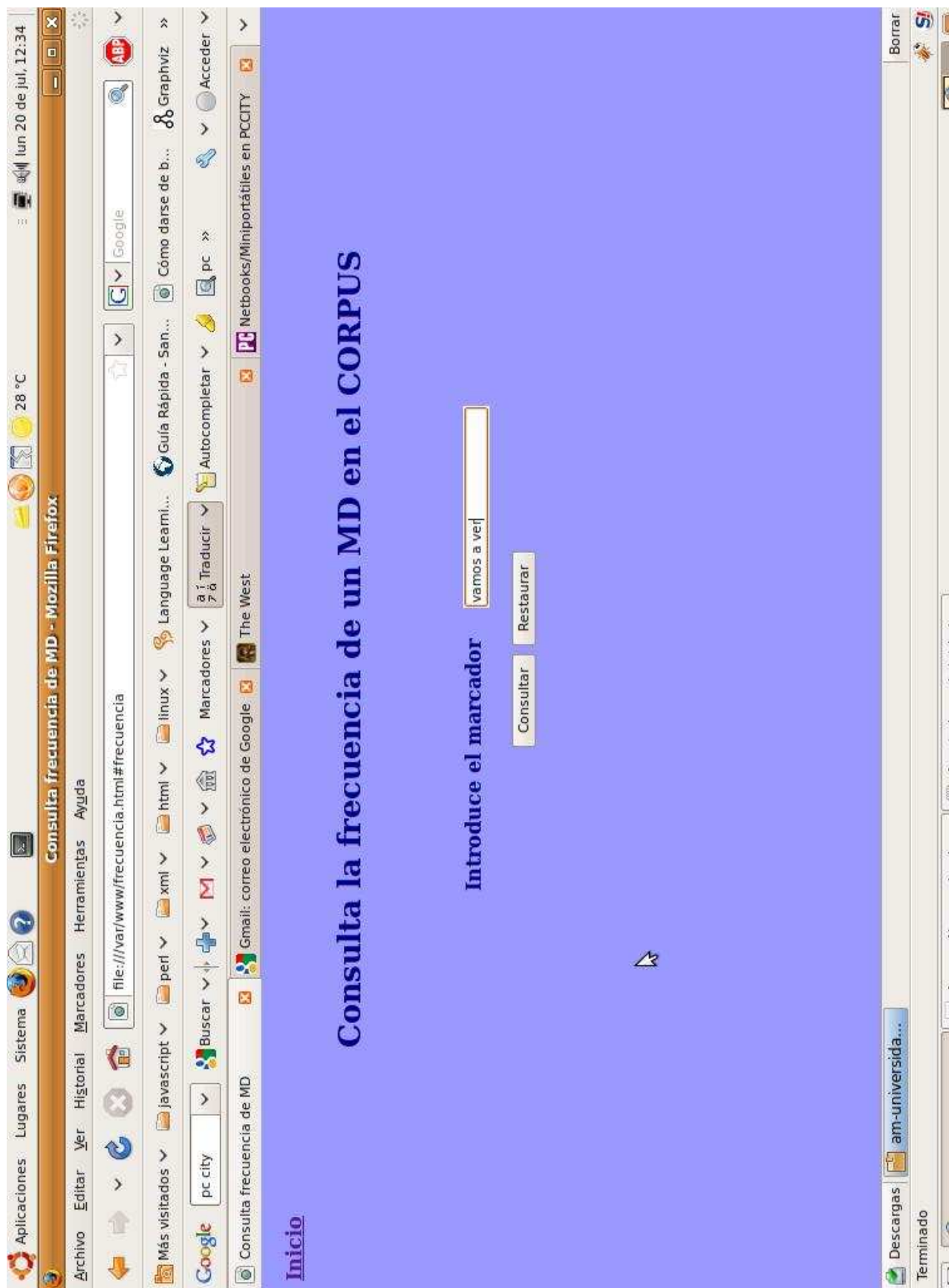


Figura 8.1: Web para lingüistas: haz click en consultar la frecuencia de un MD en el corpus



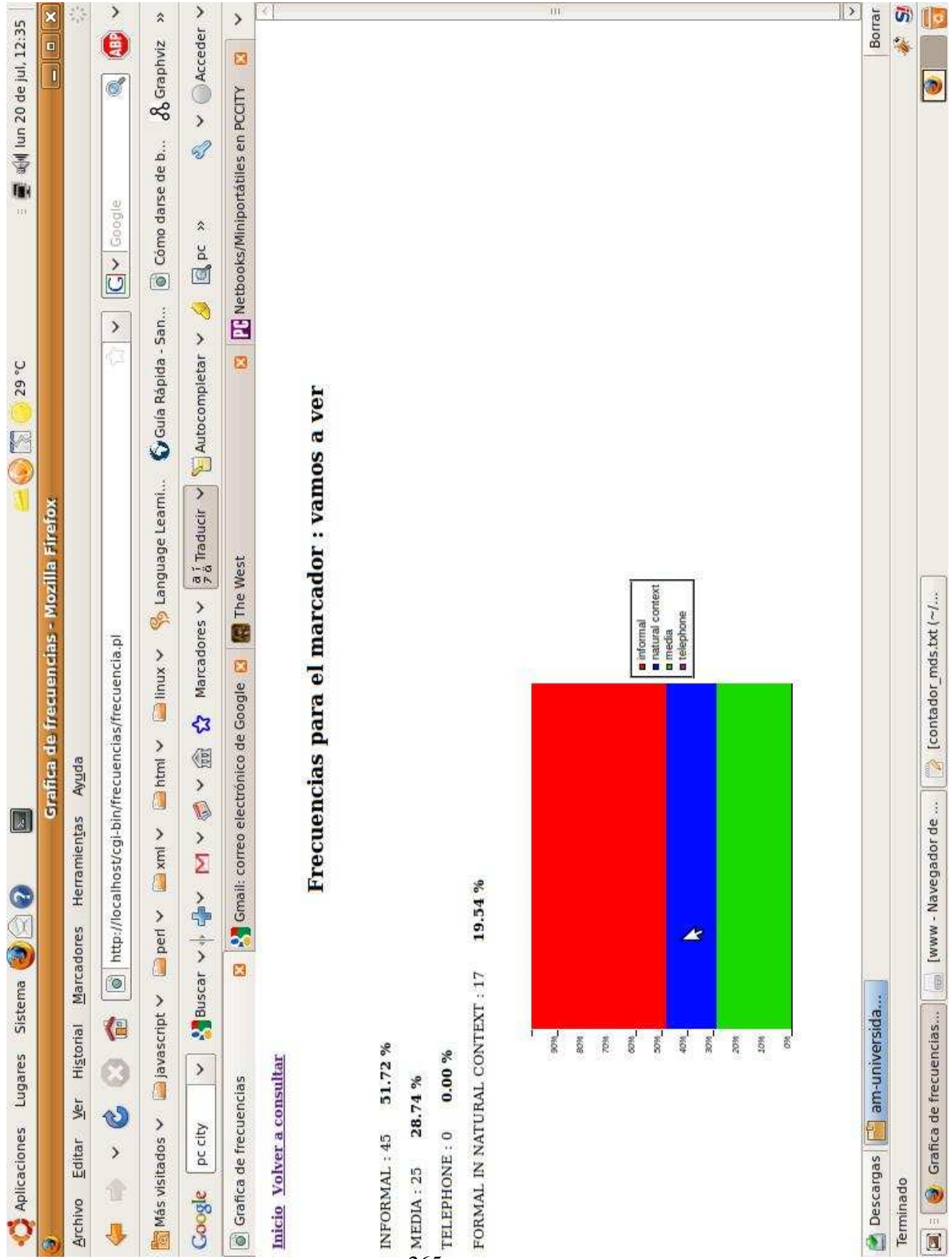


Figura 8.3: Web para lingüistas: resultados de frecuencias según el género discursivo

8.4. Frecuencia de marcadores del discurso en función de los géneros discursivos del corpus C-ORAL-ROM

A continuación se presentarán las tablas de frecuencias de marcadores. En ellas hay bastante información, con lo que comentaremos algunos aspectos sobresalientes de los resultados encontrados. Y posteriormente, en futuras publicaciones, explotaremos más detenidamente estos resultados.

Antes de pasar al análisis de los marcadores según el fenómeno pragmático al que remiten, vamos a hacer previamente una reflexión en torno a la posible relevancia de los resultados encontrados aquí. Considero de rigor señalar que a la hora de interpretar los resultados he aumentado la relevancia de las frecuencias cuando se tratan de diferencias de 0 a 1 por ejemplo, mientras que he relativizado dicha relevancia a medida que las frecuencias eran más altas. Pondré un ejemplo. Considero relevante el hecho de que un marcador del discurso aparezca en la parte informal 5 veces y en la formal, ninguna. Mientras que considero bastante menos relevante el hecho de que un marcador como *también* aparezca en informal 382 veces y en la parte formal del texto 354 veces. A mayor número de ocurrencias mayor tiene que ser la distancia entre el número de apariciones para que la considere relevante de cara a discernir qué tipo de marcadores son propios de lo escrito y cuáles son propios del registro oral. En este sentido hay que decir que en las tablas podemos encontrar número de ocurrencias de un marcador en función de cuatro géneros: **informal** y **teléfono**, y **formal in natural context** y **media**. Los dos primeros se caracterizan por ser un discurso no planificado. En cambio, las producciones orales que componen la parte de **natural in context** y **media** pueden considerarse formales, ya que tienen un texto escrito de base sobre el que van versionando oralmente. Nos referimos pues a géneros como los sermones, los programas del tiempo, las clases o las conferencias.

8.4.1. Frecuencias de marcadores de Relación Discursiva

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DR	Definite End	y punto	4	1	1	0
		TOTAL	4	1	1	0
DR	Cataphoric Cause	resulta que	17	3	5	0
		TOTAL	17	3	5	0
DR	Discourse Production	eh	630	702	513	38
DR	Discourse Production	vamos a ver	25	14	12	0
		TOTAL	655	716	525	38

Tabla 8.4: Tabla de frecuencias de MMDD de Relaciones Discursivas

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DR	Contraargumentation	ahora bien	0	2	2	0
DR	Contraargumentation	al contrario	2	1	1	0
DR	Contraargumentation	contrariamente	0	0	1	0
DR	Contraargumentation	en cambio	5	6	4	0
DR	Contraargumentation	en caso contrario	0	3	0	0
DR	Contraargumentation	más bien al contrario	0	0	1	0
DR	Contraargumentation	no obstante	0	1	3	0
DR	Contraargumentation	pero	1648	374	329	15
DR	Contraargumentation	por contra	0	0	1	0
DR	Contraargumentation	por contraste	0	1	0	0
DR	Contraargumentation	por el contrario	0	6	3	0
DR	Contraargumentation	sin embargo	5	20	14	0
DR	Contraargumentation	sino que	13	23	10	0
DR	Contraargumentation	sólo que	1	2	1	0
		TOTAL	1674	439	370	15
DR	Digression	a propósito de	0	2	0	0
DR	Digression	a todo esto	3	0	0	0
DR	Digression	por cierto	15	2	9	0
DR	Digression	por no hablar	1	0	0	0
		TOTAL	19	4	9	0
DR	Coargumentation1	ante todo	0	1	0	0
DR	Coargumentation1	antes de nada	0	1	0	0
DR	Coargumentation1	antes que nada	0	0	2	0
DR	Coargumentation1	en primer lugar	1	8	5	0
DR	Coargumentation1	para empezar	3	2	0	0
DR	Coargumentation1	por una parte	5	2	1	0
DR	Coargumentation1	primeramente	0	3	0	0
DR	Coargumentation1	de entrada	2	2	2	0
DR	Coargumentation1	primero	17	17	5	0
		TOTAL	28	36	15	0
DR	Indefinite Transitional End	etcétera	20	51	19	0
DR	Indefinite Transitional End	no sé qué historia	1	0	0	0
DR	Indefinite Transitional End	no sé qué rollo	2	0	0	0
DR	Indefinite Transitional End	o algo	47	1	3	3
DR	Indefinite Transitional End	o algo de eso	6	0	0	0
DR	Indefinite Transitional End	o algo por el estilo	1	0	0	0
DR	Indefinite Transitional End	pues nada	59	0	2	5

Tabla 8.4: Tabla de frecuencias de MMDD de Relaciones Discursivas

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DR	Indefinite Transitional End	que tal y cual	2	0	0	0
DR	Indefinite Transitional End	tal y cual	5	0	0	0
DR	Indefinite Transitional End	toda la leche	1	0	0	1
DR	Indefinite Transitional End	total que	11	0	0	0
DR	Indefinite Transitional End	y tal	112	8	6	0
DR	Indefinite Transitional End	y tal y cual	2	0	0	0
DR	Indefinite Transitional End	y toda la leche	1	0	0	0
DR	Indefinite Transitional End	y todo ese ló	1	0	0	0
DR	Indefinite Transitional End	y todo eso	21	12	0	0
DR	Indefinite Transitional End	y ya está	40	0	0	2
DR	Indefinite Transitional End	no sé cuántos	5	0	1	0
DR	Indefinite Transitional End	no sé qué no sé cuántos	6	0	0	0
DR	Indefinite Transitional End	tal	52	3	4	0
DR	Indefinite Transitional End	total	32	0	1	0
DR	Indefinite Transitional End	y demás	4	9	1	0
DR	Indefinite Transitional End	y eso	60	17	2	0
DR	Indefinite Transitional End	y nada	109	3	1	0
DR	Indefinite Transitional End	y todo	21	2	0	1
DR	Indefinite Transitional End	y ya	43	4	3	2
		TOTAL	664	110	43	14
DR	Coargumentation2	a continuación	3	2	0	0
DR	Coargumentation2	a su vez	3	4	3	0
DR	Coargumentation2	además	145	42	43	0
DR	Coargumentation2	aparte de eso	2	0	1	0
DR	Coargumentation2	aparte de que	5	1	3	0
DR	Coargumentation2	así las cosas	0	0	1	0
DR	Coargumentation2	asimismo	0	1	0	0
DR	Coargumentation2	de la misma manera	0	2	1	0
DR	Coargumentation2	de manera que	4	2	0	0
DR	Coargumentation2	de modo que	0	2	4	0
DR	Coargumentation2	en parte	0	8	1	1

Tabla 8.4: Tabla de frecuencias de MMDD de Relaciones Discursivas

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DR	Coargumentation2	en segundo lugar	0	4	2	0
DR	Coargumentation2	en tercer lugar	0	1	0	0
DR	Coargumentation2	eso por una parte	0	0	1	0
DR	Coargumentation2	por otra parte	6	5	5	0
DR	Coargumentation2	por otro lado	10	6	3	0
DR	Coargumentation2	también	376	200	154	6
DR	Coargumentation2	tampoco	146	24	18	1
DR	Coargumentation2	y bien	0	0	1	0
DR	Coargumentation2	igualmente	0	0	2	0
DR	Coargumentation2	aparte de	1	4	2	0
DR	Coargumentation2	aparte	6	4	0	0
DR	Coargumentation2	de un lado	0	1	0	0
DR	Coargumentation2	ni	178	63	38	1
DR	Coargumentation2	por otro	0	1	0	0
DR	Coargumentation2	por un lado	2	12	3	0
DR	Coargumentation2	segundo	6	4	3	0
DR	Coargumentation2	tercero	0	4	0	0
DR	Coargumentation2	y	4865	2019	1702	488
DR	Coargumentation2	que si	31	10	4	0
		TOTAL	5789	2426	1995	497
DR	Coargumentation3	en última instancia	1	0	0	0
DR	Coargumentation3	en último término	3	1	0	0
DR	Coargumentation3	por último	2	5	4	0
DR	Coargumentation3	si es que es verdad	1	0	0	0
DR	Coargumentation3	si es que eso	2	0	0	0
		TOTAL	9	6	4	0
DR	Reformulation	a ver si me explico	0	1	0	0
DR	Reformulation	dicho de otro modo	0	1	0	0
DR	Reformulation	es decir	66	157	68	0
DR	Reformulation	mejor dicho	1	1	0	0
DR	Reformulation	miento	1	0	0	0
DR	Reformulation	o lo que es lo mismo	0	1	0	0
DR	Reformulation	o sea	556	86	56	8
DR	Reformulation	por así decirlo	1	0	0	0
DR	Reformulation	qué digo	1	0	0	0
DR	Reformulation	quiero decir que	3	2	3	0
DR	Reformulation	si me explico	1	0	0	0
DR	Reformulation	digamos	16	30	6	0
DR	Reformulation	esto es	10	20	9	0
DR	Reformulation	quiero decir	0	8	4	0
DR	Reformulation	vamos	105	61	20	0

Tabla 8.4: Tabla de frecuencias de MMDD de Relaciones Discursivas

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DR	Reformulation	a saber	3	1	0	0
		TOTAL	764	369	166	8
DR	Hypothesis	a lo mejor	110	17	11	2
DR	Hypothesis	a lo peor	0	0	1	0
DR	Hypothesis	a menos que	0	1	0	0
DR	Hypothesis	a no ser que	2	0	1	0
DR	Hypothesis	a ver si	58	7	6	3
DR	Hypothesis	acaso	0	1	1	0
DR	Hypothesis	como si	19	10	6	0
DR	Hypothesis	con tal de que	2	0	0	0
DR	Hypothesis	mira que si	1	0	0	0
DR	Hypothesis	por si acaso	8	1	0	0
DR	Hypothesis	si	881	288	141	41
DR	Hypothesis	si acaso	1	3	3	1
DR	Hypothesis	siempre y cuando	1	0	0	0
DR	Hypothesis	como	12	1	0	0
		TOTAL	1095	329	170	47
DR	Option	no ?	417	167	181	6
DR	Option	o	597	387	225	115
DR	Option	o bien	12	9	8	7
DR	Option	o bien que	0	1	0	0
DR	Option	o no ?	27	2	3	0
DR	Option	o qué ?	12	1	0	0
DR	Option	o sí ?	0	0	1	0
		TOTAL	1065	567	418	128
DR	Conclusion Summary	a tomar por culo	3	0	0	0
DR	Conclusion Summary	al fin y al cabo	3	0	2	0
DR	Conclusion Summary	al final	90	12	12	1
DR	Conclusion Summary	en conclusión	0	1	0	0
DR	Conclusion Summary	en una palabra	0	1	0	0
DR	Conclusion Summary	fundamentalmente	0	7	13	0
DR	Conclusion Summary	grosso modo	0	1	0	0
DR	Conclusion Summary	principalmente	1	8	3	0
DR	Conclusion Summary	resumiendo un poco	0	0	1	0
DR	Conclusion Summary	vamos a decirlo muy brevemente	0	1	0	0
DR	Conclusion Summary	conclusión	0	2	1	0
DR	Conclusion Summary	en fin	28	17	11	0
DR	Conclusion Summary	en total	1	0	0	0
		TOTAL	126	50	43	1
DR	Cause	dado que	0	2	1	0
DR	Cause	debido a que	0	2	1	0

Tabla 8.4: Tabla de frecuencias de MMDD de Relaciones Discursivas

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DR	Cause	es que	395	86	26	2
DR	Cause	lo que pasa	3	3	1	0
DR	Cause	por eso mismo	0	2	2	0
DR	Cause	por eso te digo	6	1	1	3
DR	Cause	por este motivo	0	1	1	0
DR	Cause	porque	1129	322	223	33
DR	Cause	puesto que	3	7	1	0
DR	Cause	y no por ello	0	0	1	0
DR	Cause	como	214	37	17	4
DR	Cause	gracias a	1	1	1	0
DR	Cause	ya que	4	4	6	0
		TOTAL	1755	557	282	42
DR	Consequence	así que	50	15	6	6
DR	Consequence	así que ...	19	0	0	2
DR	Consequence	así que bueno	2	0	1	1
DR	Consequence	así que mira	2	0	0	0
DR	Consequence	así que nada	5	0	0	0
DR	Consequence	así que no sé	1	0	0	0
DR	Consequence	así que por eso te digo	0	0	0	1
DR	Consequence	con lo cual	19	31	1	0
DR	Consequence	en consecuencia	1	0	2	0
DR	Consequence	entonces	626	217	72	13
DR	Consequence	entonces ?	8	1	0	0
DR	Consequence	por lo cual	0	0	2	0
DR	Consequence	por lo tanto	12	42	22	0
DR	Consequence	por tanto	2	18	15	0
DR	Consequence	pues	1368	388	230	36
		TOTAL	2115	712	351	59
DR	Anaphoric Concessiveness Cataphoric Conclusion	a fin de cuentas	1	2	1	0
DR	Anaphoric Concessiveness Cataphoric Conclusion	de todas formas	17	4	2	2
DR	Anaphoric Concessiveness Cataphoric Conclusion	de todas maneras	16	3	4	1
DR	Anaphoric Concessiveness Cataphoric Conclusion	de todos modos	3	0	1	0
DR	Anaphoric Concessiveness Cataphoric Conclusion	en cualquier caso	6	4	3	0

Tabla 8.4: Tabla de frecuencias de MMDD de Relaciones Discursivas

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DR	Anaphoric Conces- siveness Cataphoric- Conclusion	en todo caso	1	6	5	0
DR	Anaphoric Conces- siveness Cataphoric- Conclusion	con todo	6	1	1	0
		TOTAL	50	20	17	3
DR	Framework	desde el punto de vista	4	0	2	0
DR	Framework	en tanto que	0	2	0	0
DR	Framework	desde el punto de vista de	0	1	3	0
DR	Framework	desde mi punto de vista	1	0	0	0
DR	Framework	desde nuestro punto de vista	1	1	0	0
DR	Framework	en el sentido de que	0	7	0	0
DR	Framework	en este sentido	0	13	4	0
DR	Framework	en mi opinión	0	1	1	0
DR	Framework	de acuerdo con	0	1	1	0
		TOTAL	6	26	11	0
DR	Generalization	en general	3	24	7	0
DR	Generalization	eso es como todo	1	0	0	0
DR	Generalization	normalmente	19	30	5	0
DR	Generalization	por lo general	1	1	1	0
		TOTAL	24	55	13	0
DR	Concretness	el caso es que	9	0	3	0
DR	Concretness	en concreto	2	3	2	0
DR	Concretness	en el caso de	4	11	4	0
DR	Concretness	en el caso del	0	1	1	0
DR	Concretness	en ese caso	0	1	0	0
DR	Concretness	la cosa es que	2	0	0	0
DR	Concretness	la cuestión es que	7	0	0	0
DR	Concretness	lo que pasa es que	29	10	5	0
DR	Concretness	lo que pasa que	32	4	1	1
DR	Concretness	mismamente	1	2	0	0
DR	Concretness	por ejemplo	117	83	51	7
DR	Concretness	concretamente	1	4	1	0
		TOTAL	204	119	68	8
DR	Anaphoric Cause Cat- aphoricConsequence	de ahí que	0	1	0	0
DR	Anaphoric Cause Cat- aphoricConsequence	de ahí	12	7	2	0
DR	Anaphoric Cause Cat- aphoricConsequence	de aquí	7	6	1	0

Tabla 8.4: Tabla de frecuencias de MMDD de Relaciones Discursivas

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DR	Anaphoric Cause CataphoricConsequence	por ello	0	2	0	0
DR	Anaphoric Cause CataphoricConsequence	por eso	31	18	6	1
		TOTAL	50	34	9	1
DR	Scalar Meaning	aún	8	6	6	0
DR	Scalar Meaning	inclusive	2	1	0	0
DR	Scalar Meaning	incluso	32	45	27	0
DR	Scalar Meaning	más aún	0	1	0	0
DR	Scalar Meaning	ni siquiera	15	6	5	0
DR	Scalar Meaning	casi	57	19	38	0
DR	Scalar Meaning	encima	30	6	4	0
		TOTAL	144	82	80	0
		TOTAL	16261	6575	4593	861

Tabla 8.4: Tabla de frecuencias de MMDD de Relaciones Discursivas

8.4.2. Frecuencias de marcadores de Evaluación Discursiva

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
ED	Positive Evaluation	¡ja!	8	0	0	0
ED	Positive Evaluation	¡ji!	2	0	0	0
ED	Positive Evaluation	¡jo!	43	1	0	0
ED	Positive Evaluation	¡ole!	0	0	1	0
ED	Positive Evaluation	¡yeah!	2	0	0	0
ED	Positive Evaluation	a lo mejor	110	17	11	2
ED	Positive Evaluation	al final	90	12	12	1
ED	Positive Evaluation	alucinante	4	0	0	0
ED	Positive Evaluation	bien es cierto	0	1	2	0
ED	Positive Evaluation	bien es cierto que	0	0	1	0
ED	Positive Evaluation	es chévere	1	0	0	0
ED	Positive Evaluation	es maravilloso	1	0	0	0
ED	Positive Evaluation	eso es alucinante	2	0	0	0
ED	Positive Evaluation	eso es chévere	1	0	0	0

Tabla 8.5: Tabla de frecuencias de MMDD de Evaluación Emocional

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
ED	Positive Evaluation	eso es la leche	1	0	0	0
ED	Positive Evaluation	eso está bien	3	0	1	0
ED	Positive Evaluation	eso está guay	2	0	0	0
ED	Positive Evaluation	está bien	16	2	0	2
ED	Positive Evaluation	está de puta madre	1	0	0	0
ED	Positive Evaluation	está guay	2	0	0	0
ED	Positive Evaluation	está muy guay	1	0	0	0
ED	Positive Evaluation	gracias a Dios	2	1	0	0
ED	Positive Evaluation	la mar de bien	0	0	1	0
ED	Positive Evaluation	la verdad es que está muy guay	1	0	0	0
ED	Positive Evaluation	más bien que	0	1	0	0
ED	Positive Evaluation	me parece estupendo	1	0	0	0
ED	Positive Evaluation	menos mal	4	0	0	0
ED	Positive Evaluation	menos mal que	5	0	0	0
ED	Positive Evaluation	por fin	2	0	1	0
ED	Positive Evaluation	por fortuna	0	1	0	0
ED	Positive Evaluation	pues qué bien	1	0	0	0
ED	Positive Evaluation	qué gracia	2	0	0	0
ED	Positive Evaluation	qué guay	8	0	0	0
ED	Positive Evaluation	todo guay	1	0	0	0
ED	Positive Evaluation	y bien	0	0	1	0
ED	Positive Evaluation	¡joder!	5	0	0	0
ED	Positive Evaluation	es espectacular	0	0	2	0
ED	Positive Evaluation	espectacular	0	0	2	0
ED	Positive Evaluation	estupendo	9	1	0	0
ED	Positive Evaluation	genial	9	0	0	0

Tabla 8.5: Tabla de frecuencias de MMDD de Evaluación Emocional

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
ED	Positive Evalua- tion	gracias a	1	1	1	0
ED	Positive Evalua- tion	guay	2	0	0	0
ED	Positive Evalua- tion	la hostia	2	0	0	0
ED	Positive Evalua- tion	maravilloso	0	1	0	0
ED	Positive Evalua- tion	muy bien	107	8	14	1
ED	Positive Evalua- tion	qué bien	7	0	1	0
		TOTAL	459	47	51	6
ED	Negative Evalua- tion	¡ay madre!	1	0	0	0
ED	Negative Evalua- tion	¡ay!	63	8	4	0
ED	Negative Evalua- tion	¡bah!	7	0	0	0
ED	Negative Evalua- tion	¡buah!	6	0	1	0
ED	Negative Evalua- tion	¡cachis en la mar!	1	0	0	0
ED	Negative Evalua- tion	¡Dios mío de mi alma!	1	0	0	0
ED	Negative Evalua- tion	¡hala!	17	0	0	0
ED	Negative Evalua- tion	¡huy!	18	1	5	0
ED	Negative Evalua- tion	¡jobar!	4	0	0	0
ED	Negative Evalua- tion	¡jolín!	3	0	0	1
ED	Negative Evalua- tion	¡jelines!	2	0	0	0
ED	Negative Evalua- tion	¡jope!	1	0	0	0
ED	Negative Evalua- tion	¡leche!	2	0	0	0
ED	Negative Evalua- tion	¡por Dios!	3	0	0	0
ED	Negative Evalua- tion	¡uf!	1	0	0	0
ED	Negative Evalua- tion	¡ups!	2	0	0	0
ED	Negative Evalua- tion	¡uy!	7	1	1	0
ED	Negative Evalua- tion	¡va!	0	1	0	0
ED	Negative Evalua- tion	¡vamos!	1	0	0	0
ED	Negative Evalua- tion	a lo peor	0	0	1	0

Tabla 8.5: Tabla de frecuencias de MMDD de Evaluación Emocional

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
ED	Negative Evaluation	a pesar	0	1	1	0
ED	Negative Evaluation	a pesar de	1	2	6	0
ED	Negative Evaluation	a pesar de que	1	3	1	0
ED	Negative Evaluation	a tomar por culo	3	0	0	0
ED	Negative Evaluation	bah	1	0	0	0
ED	Negative Evaluation	coño	1	0	0	0
ED	Negative Evaluation	cuando	524	156	125	8
ED	Negative Evaluation	es lo peor	1	0	0	0
ED	Negative Evaluation	hala	1	0	0	0
ED	Negative Evaluation	hasta los cojones	2	0	0	0
ED	Negative Evaluation	jolín	2	0	0	0
ED	Negative Evaluation	la verdad es que no mola mucho	1	0	0	0
ED	Negative Evaluation	me cago en diona	1	0	0	0
ED	Negative Evaluation	me cago en Dios	1	0	0	0
ED	Negative Evaluation	no me jodas	1	0	0	0
ED	Negative Evaluation	no te jode	3	0	0	0
ED	Negative Evaluation	pese a que	1	0	0	0
ED	Negative Evaluation	por desgracia	0	1	3	0
ED	Negative Evaluation	qué coño mal	1	0	0	0
ED	Negative Evaluation	que le den por culo	1	0	0	0
ED	Negative Evaluation	que no veas	8	1	0	0
ED	Negative Evaluation	joder	9	0	0	0
ED	Negative Evaluation	coño	1	0	0	0
ED	Negative Evaluation	¡joder!	21	0	0	0
ED	Negative Evaluation	encima	30	6	4	0
ED	Negative Evaluation	por Dios	0	1	0	0

Tabla 8.5: Tabla de frecuencias de MMDD de Evaluación Emocional

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
ED	Negative Evaluation	una mierda	2	0	0	0
ED	Negative Evaluation	vaya	17	6	1	0
ED	Negative Evaluation	que si	31	10	4	0
		TOTAL	806	198	157	9
ED	Surprise	¡ah!	404	11	17	22
ED	Surprise	¡ahí va!	2	0	0	0
ED	Surprise	¡hostia!	6	0	0	0
ED	Surprise	¡hostias!	2	0	0	0
ED	Surprise	¡madre mía!	16	0	1	1
ED	Surprise	¡oh!	9	0	1	0
ED	Surprise	¡ostras!	6	0	0	1
ED	Surprise	ah	18	13	67	1
ED	Surprise	de la hostia	1	0	0	0
ED	Surprise	eso es muy fuerte	1	0	0	0
ED	Surprise	esto ya es muy fuerte	1	0	0	0
ED	Surprise	madre mía	12	0	0	0
ED	Surprise	no me lo puedo creer	1	0	0	0
ED	Surprise	ostras	1	0	0	0
ED	Surprise	qué barbaridad	2	0	0	1
ED	Surprise	qué fuerte	7	0	0	0
ED	Surprise	¡hombre!	1	0	0	0
ED	Surprise	hombre	1	0	0	0
ED	Surprise	es muy fuerte	0	1	0	0
ED	Surprise	fuerte	5	4	4	0
ED	Surprise	hostia	1	0	0	0
ED	Surprise	muy fuerte	2	1	0	0
		TOTAL	499	30	90	26
		TOTAL	1764	275	298	41

Tabla 8.5: Tabla de frecuencias de MMDD de Evaluación Emocional

8.4.3. Frecuencias de marcadores de Modalización discursiva

8.4.3.1. Frecuencias de marcadores de intensificación, atenuación e interacción

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
MOD	DR Hedge	a lo mejor	110	17	11	2
MOD	DR Hedge	a lo peor	0	0	1	0
MOD	DR Hedge	a menos que	0	1	0	0

Tabla 8.6: Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
MOD	DR Hedge	a no ser que	2	0	1	0
MOD	DR Hedge	así que bueno	2	0	1	1
MOD	DR Hedge	así que nada	5	0	0	0
MOD	DR Hedge	así que no sé	1	0	0	0
MOD	DR Hedge	como si	19	10	6	0
		TOTAL	139	28	20	3
MOD	DR Booster	así que por eso te digo	0	0	0	1
MOD	DR Booster	es que no veas	1	0	0	0
MOD	DR Booster	no veas	6	0	0	1
MOD	DR Booster	que no veas	8	1	0	0
MOD	DR Booster	que te cagas	10	0	0	0
MOD	DR Booster	justo	4	1	0	0
		TOTAL	29	2	0	2
MOD	Interactional Hedge	venga	36	3	7	0
		TOTAL	36	3	7	0
MOD	Acte Hedge	hija mía	2	0	1	0
MOD	Acte Hedge	no por otra cosa	1	0	0	0
MOD	Acte Hedge	no te digo nada	2	0	0	0
MOD	Acte Hedge	no te digo que no	2	0	0	0
MOD	Acte Hedge	no te jode	3	0	0	0
MOD	Acte Hedge	por favor	14	4	3	582
MOD	Acte Hedge	porfa	2	0	0	0
MOD	Acte Hedge	perdón	10	11	5	0
MOD	Acte Hedge	qué demonios	0	0	1	0
		TOTAL	36	15	10	582
MOD	Formulation Hedge	no sé cuántos	5	0	1	0
MOD	Formulation Hedge	no sé qué no sé cuántos	6	0	0	0
MOD	Formulation Hedge	y eso	60	17	2	0
MOD	Formulation Hedge	y nada	109	3	1	0
MOD	Formulation Hedge	y todo	21	2	0	1
MOD	Formulation Hedge	y ya	43	4	3	2
		TOTAL	244	26	7	3
MOD	Truth Booster	¡vamos!	1	0	0	0
MOD	Truth Booster	a decir verdad	1	0	0	0
MOD	Truth Booster	absolutamente	7	15	13	0
MOD	Truth Booster	casi seguro	3	1	0	0

Tabla 8.6: Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
MOD	Truth Booster	ciertamente	1	1	8	0
MOD	Truth Booster	claramente	0	7	6	0
MOD	Truth Booster	claro que no	3	0	0	0
MOD	Truth Booster	claro que sí	3	0	4	0
MOD	Truth Booster	como es natural	2	0	3	0
MOD	Truth Booster	con toda seguridad	0	0	1	0
MOD	Truth Booster	de hecho	22	12	10	1
MOD	Truth Booster	desde luego	15	16	10	0
MOD	Truth Booster	efectivamente	31	19	21	2
MOD	Truth Booster	en absoluto	1	6	7	0
MOD	Truth Booster	en efecto	0	1	1	0
MOD	Truth Booster	en el fondo	4	2	3	0
MOD	Truth Booster	en realidad	11	12	6	1
MOD	Truth Booster	es cierto	3	6	3	0
MOD	Truth Booster	es cierto que	1	1	0	0
MOD	Truth Booster	es evidente que	0	0	2	0
MOD	Truth Booster	es muy probable	3	0	0	0
MOD	Truth Booster	es obvio	1	0	0	0
MOD	Truth Booster	es verdad	43	5	6	0
MOD	Truth Booster	es verdad que	0	4	3	0
MOD	Truth Booster	eso es cierto	1	0	0	0
MOD	Truth Booster	eso es evidente	0	0	1	0
MOD	Truth Booster	eso es obvio	1	0	0	0
MOD	Truth Booster	eso está claro	2	0	1	0
MOD	Truth Booster	eso está muy claro	1	0	0	0
MOD	Truth Booster	eso mismo	1	0	0	0
MOD	Truth Booster	eso no es cierto	0	1	0	0
MOD	Truth Booster	eso sí	16	0	1	0
MOD	Truth Booster	eso sí que	6	0	1	0
MOD	Truth Booster	eso también está su- perclaro	1	0	0	0
MOD	Truth Booster	especialmente	7	2	6	0
MOD	Truth Booster	está claro que	1	0	0	0
MOD	Truth Booster	evidentemente	5	19	10	0
MOD	Truth Booster	indefectiblemente	0	1	0	0
MOD	Truth Booster	indudablemente	0	0	10	0
MOD	Truth Booster	inevitablemente	0	0	1	0
MOD	Truth Booster	justamente	3	6	8	0
MOD	Truth Booster	la verdad es que	34	4	13	0
MOD	Truth Booster	la verdad es que es eso	1	0	0	0
MOD	Truth Booster	la verdad es que está muy guay	1	0	0	0

Tabla 8.6: Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
MOD	Truth Booster	la verdad es que no	2	0	0	0
MOD	Truth Booster	la verdad es que no mo- la mucho	1	0	0	0
MOD	Truth Booster	la verdad es que sí	3	0	0	0
MOD	Truth Booster	lo cierto es que	0	1	1	0
MOD	Truth Booster	lo que está claro es que	1	1	0	0
MOD	Truth Booster	lo que sí es cierto	0	3	0	0
MOD	Truth Booster	lógicamente	1	13	4	0
MOD	Truth Booster	muy claramente	0	0	1	0
MOD	Truth Booster	nada en absoluto	1	0	0	0
MOD	Truth Booster	naturalmente	1	1	5	0
MOD	Truth Booster	necesariamente	0	4	2	0
MOD	Truth Booster	ni nada	35	4	2	0
MOD	Truth Booster	ni nada de eso	3	0	0	0
MOD	Truth Booster	obviamente	4	4	5	0
MOD	Truth Booster	obviamente no	1	0	0	0
MOD	Truth Booster	por eso mismo	0	2	2	0
MOD	Truth Booster	por eso te digo	6	1	1	3
MOD	Truth Booster	por fuerza	0	1	0	0
MOD	Truth Booster	por los cuatro costados	0	0	1	0
MOD	Truth Booster	por supuesto	24	12	8	0
MOD	Truth Booster	por supuesto que sí	0	0	1	0
MOD	Truth Booster	prácticamente	12	16	15	0
MOD	Truth Booster	precisamente	4	11	18	0
MOD	Truth Booster	presuntamente	0	0	1	0
MOD	Truth Booster	pues eso mismo	1	0	0	0
MOD	Truth Booster	pues sí que	7	0	0	0
MOD	Truth Booster	qué duda cabe	0	2	0	0
MOD	Truth Booster	queda claro ?	1	0	0	0
MOD	Truth Booster	queda claro todo esto ?	1	0	0	0
MOD	Truth Booster	realmente	36	20	21	0
MOD	Truth Booster	se ve que	7	1	0	0
MOD	Truth Booster	seguramente	8	8	5	0
MOD	Truth Booster	seguramente sí	2	0	0	0
MOD	Truth Booster	seguro que	25	2	0	0
MOD	Truth Booster	si bien es verdad que	1	0	0	0
MOD	Truth Booster	sí es cierto que	1	0	0	0
MOD	Truth Booster	si es que es verdad	1	0	0	0
MOD	Truth Booster	si es que eso	2	0	0	0
MOD	Truth Booster	sí es verdad	2	0	0	0
MOD	Truth Booster	sí que es cierto	0	1	0	0
MOD	Truth Booster	sí que es verdad	0	0	1	0

Tabla 8.6: Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
MOD	Truth Booster	si te digo la verdad	1	0	0	0
MOD	Truth Booster	siempre	180	51	58	2
MOD	Truth Booster	siempre que	13	2	0	0
MOD	Truth Booster	sin duda	0	2	5	0
MOD	Truth Booster	también es verdad	2	1	1	0
MOD	Truth Booster	ya ves	9	1	0	0
MOD	Truth Booster	ya ves que	1	0	0	0
MOD	Truth Booster	así es	1	3	1	0
MOD	Truth Booster	claro	258	34	15	3
MOD	Truth Booster	de acuerdo con	0	1	1	0
MOD	Truth Booster	de acuerdo	7	3	5	1
MOD	Truth Booster	de corazón	0	1	1	0
MOD	Truth Booster	de verdad	7	5	2	0
MOD	Truth Booster	evidente	2	3	1	0
MOD	Truth Booster	exactamente	8	7	2	0
MOD	Truth Booster	la verdad	13	9	8	0
MOD	Truth Booster	ni	178	63	38	1
MOD	Truth Booster	pues eso	35	4	2	0
MOD	Truth Booster	seguro	5	2	0	0
MOD	Truth Booster	vale	58	8	1	7
MOD	Truth Booster	ya te digo	5	2	0	0
MOD	Truth Booster	para nada	2	0	0	0
MOD	Truth Booster	particularmente	1	1	0	0
		TOTAL	1217	452	394	21
MOD	Truth Hedge	¡anda!	19	2	2	0
MOD	Truth Hedge	¡bah!	7	0	0	0
MOD	Truth Hedge	¡buah!	6	0	1	0
MOD	Truth Hedge	a mi modo de ver	0	1	0	0
MOD	Truth Hedge	aproximadamente	1	3	2	1
MOD	Truth Hedge	bah	1	0	0	0
MOD	Truth Hedge	como que	17	3	3	0
MOD	Truth Hedge	desde mi punto de vista	1	0	0	0
MOD	Truth Hedge	desde nuestro punto de vista	1	1	0	0
MOD	Truth Hedge	en mi opinión	0	1	1	0
MOD	Truth Hedge	en parte	0	8	1	1
MOD	Truth Hedge	entre comillas	2	1	2	0
MOD	Truth Hedge	entre paréntesis	0	1	0	0
MOD	Truth Hedge	es casi seguro que	0	0	1	0
MOD	Truth Hedge	más bien	6	11	4	0
MOD	Truth Hedge	más bien al contrario	0	0	1	0
MOD	Truth Hedge	más bien que	0	1	0	0

Tabla 8.6: Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
MOD	Truth Hedge	más o menos	47	25	12	2
MOD	Truth Hedge	miento	1	0	0	0
MOD	Truth Hedge	no ?	417	167	181	6
MOD	Truth Hedge	o algo	47	1	3	3
MOD	Truth Hedge	o algo ?	3	0	0	0
MOD	Truth Hedge	o algo de eso	6	0	0	0
MOD	Truth Hedge	o algo por el estilo	1	0	0	0
MOD	Truth Hedge	o cómo ?	3	0	0	0
MOD	Truth Hedge	o no ?	27	2	3	0
MOD	Truth Hedge	o qué ?	12	1	0	0
MOD	Truth Hedge	o sí ?	0	0	1	0
MOD	Truth Hedge	por así decirlo	1	0	0	0
MOD	Truth Hedge	por lo visto	5	1	0	0
MOD	Truth Hedge	posiblemente	2	2	5	0
MOD	Truth Hedge	probablemente	11	12	9	0
MOD	Truth Hedge	qué coño mal	1	0	0	0
MOD	Truth Hedge	qué digo	1	0	0	0
MOD	Truth Hedge	que recuerde	0	0	1	0
MOD	Truth Hedge	que yo sepa	2	0	1	0
MOD	Truth Hedge	quiero decir que	3	2	3	0
MOD	Truth Hedge	quizá	5	11	16	0
MOD	Truth Hedge	quizás	8	6	0	0
MOD	Truth Hedge	relativamente	1	3	3	0
MOD	Truth Hedge	se supone	7	2	1	0
MOD	Truth Hedge	se supone que	9	5	0	0
MOD	Truth Hedge	según parece	0	0	1	0
MOD	Truth Hedge	si	881	288	141	41
MOD	Truth Hedge	si acaso	1	3	3	1
MOD	Truth Hedge	supongo	7	0	0	0
MOD	Truth Hedge	supongo que	22	2	2	0
MOD	Truth Hedge	supongo que sí	0	0	1	0
MOD	Truth Hedge	supongo yo	1	0	0	0
MOD	Truth Hedge	supuestamente	1	2	1	0
MOD	Truth Hedge	tal vez	2	1	2	0
MOD	Truth Hedge	vete a saber	1	1	2	0
MOD	Truth Hedge	ya ves tú	9	0	0	0
MOD	Truth Hedge	¡hombre!	12	0	0	0
MOD	Truth Hedge	hombre	40	0	12	0
MOD	Truth Hedge	anda	5	0	0	0
MOD	Truth Hedge	bueno	519	130	78	8
MOD	Truth Hedge	digamos	16	30	6	0

Tabla 8.6: Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
MOD	Truth Hedge	lo que se dice	1	0	1	0
MOD	Truth Hedge	que digamos	1	2	0	0
MOD	Truth Hedge	quiero decir	0	8	4	0
MOD	Truth Hedge	vamos	105	61	20	0
		TOTAL	2305	801	531	63
MOD	Hedge Action	cariño	4	0	2	0
		TOTAL	4	0	2	0
MOD	Surprise Booster	¡madre mía!	16	0	1	1
MOD	Surprise Booster	de la hostia	1	0	0	0
MOD	Surprise Booster	no me lo puedo creer	1	0	0	0
MOD	Surprise Booster	qué barbaridad	2	0	0	1
		TOTAL	20	0	1	2
		TOTAL	0	0	0	0
MOD	Acte Booster	¡ajá!	8	0	1	0
MOD	Acte Booster	de eso nada	2	0	0	0
MOD	Acte Booster	faltaría más	1	0	0	0
MOD	Acte Booster	nanai	1	0	0	0
MOD	Acte Booster	naranjas de la China	1	0	0	0
MOD	Acte Booster	ni de coña	6	0	0	0
		TOTAL	19	0	1	0
MOD	Emotion Booster	¡ole!	0	0	1	0
MOD	Emotion Booster	bien es cierto	0	1	2	0
MOD	Emotion Booster	bien es cierto que	0	0	1	0
MOD	Emotion Booster	esto ya es muy fuerte	1	0	0	0
MOD	Emotion Booster	la mar de bien	0	0	1	0
MOD	Emotion Booster	la hostia	2	0	0	0
		TOTAL	3	1	5	0
MOD	Emotion Hedge	¡jobar!	4	0	0	0
MOD	Emotion Hedge	¡jolín!	3	0	0	1
MOD	Emotion Hedge	¡jelines!	2	0	0	0
MOD	Emotion Hedge	¡jope!	1	0	0	0
MOD	Emotion Hedge	¡leche!	2	0	0	0
MOD	Emotion Hedge	jolín	2	0	0	0
		TOTAL	14	0	0	1
MOD	Interaction	¡ah! sí ?	5	0	0	0
MOD	Interaction	¡eh!	5	0	0	0
MOD	Interaction	¡ey!	4	0	0	0
MOD	Interaction	a que sí ?	8	0	0	0
MOD	Interaction	ah ?	9	0	0	0
MOD	Interaction	cómo ?	6	0	0	1
MOD	Interaction	cómo lo ves ?	1	0	0	0

Tabla 8.6: Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
MOD	Interaction	cómo que no ?	2	1	0	0
MOD	Interaction	de verdad ?	1	0	0	0
MOD	Interaction	eh ?	219	154	56	4
MOD	Interaction	entiendes ?	1	0	0	0
MOD	Interaction	entonces ?	8	1	0	0
MOD	Interaction	lo ves ?	3	0	0	0
MOD	Interaction	me entiendes ?	5	0	0	1
MOD	Interaction	me entiendes lo que te quiero decir ?	1	0	0	0
MOD	Interaction	no te acuerdas ?	5	0	0	0
MOD	Interaction	no ves ?	1	1	0	0
MOD	Interaction	os acordáis ?	4	0	0	0
MOD	Interaction	perdona ?	2	0	0	0
MOD	Interaction	qué creéis ?	1	0	0	0
MOD	Interaction	qué dices ?	3	0	0	0
MOD	Interaction	qué pasa ?	26	3	1	1
MOD	Interaction	qué te iba a decir ?	5	0	0	0
MOD	Interaction	sabes ?	167	2	0	1
MOD	Interaction	seguro ?	3	0	0	0
MOD	Interaction	sí ?	90	3	5	3
MOD	Interaction	sí o no ?	5	0	0	0
MOD	Interaction	te acuerdas ?	3	0	0	0
MOD	Interaction	vale ?	32	5	1	2
MOD	Interaction	verdad hijo ?	1	0	0	0
MOD	Interaction	ves ?	27	0	1	0
MOD	Interaction	ves cómo es cierto ?	1	0	0	0
MOD	Interaction	y qué pasa ?	2	1	0	0
MOD	Interaction	y ya está ?	1	0	0	0
MOD	Interaction	igualmente	1	1	2	0
MOD	Interaction	escucha	3	5	1	0
MOD	Interaction	mira	76	13	1	2
MOD	Interaction	ojo	1	0	1	0
MOD	Interaction	verdad ?	6	4	2	0
		TOTAL	744	194	71	15
		TOTAL	4810	1522	1049	692

Tabla 8.6: Tabla de frecuencias de MMDD de Modalización discursiva: atenuación, intensificación e interacción

8.4.4. Frecuencia de marcadores de Evidencialidad

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
EVI	Inference	eso es como todo	1	0	0	0
EVI	Inference	lógicamente	1	13	4	0
EVI	Inference	normalmente	19	30	5	0
EVI	Inference	por fuerza	0	1	0	0
EVI	Inference	por lo general	1	1	1	0
EVI	Inference	se ve que	7	1	0	0
EVI	Inference	siempre	180	51	58	2
EVI	Inference	siempre que	13	2	0	0
		TOTAL	222	99	68	2
EVI	Cultural Knowledge	como es natural	2	0	3	0
		TOTAL	2	0	3	0
EVI	First Person	desde mi punto de vista	1	0	0	0
EVI	First Person	desde nuestro punto de vista	1	1	0	0
EVI	First Person	en mi opinión	0	1	1	0
EVI	First Person	supongo	7	0	0	0
EVI	First Person	supongo que	22	2	2	0
EVI	First Person	supongo que sí	0	0	1	0
EVI	First Person	supongo yo	1	0	0	0
EVI	First Person	supuestamente	1	2	1	0
		TOTAL	33	6	5	0
EVI	Others	por lo visto	5	1	0	0
EVI	Others	se supone	7	2	1	0
EVI	Others	se supone que	9	5	0	0
EVI	Others	según parece	0	0	1	0
EVI	Others	dicen que	2	0	0	0
EVI	Others	lo que se dice	1	0	1	0
		TOTAL	24	8	3	0
		TOTAL	281	113	79	2

Tabla 8.7: Tabla de frecuencias de MMDD de Evidencialidad

8.4.5. Frecuencia de marcadores de Actos de Habla

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
SA	Commitment	verás	6	0	0	0
		TOTAL	6	0	0	0
SA	Wish	que en paz descanse	2	0	0	0
		TOTAL	2	0	0	0
SA	Rejection	no me jodas	1	0	0	0
SA	Rejection	no te jode	3	0	0	0
SA	Rejection	que le den por culo	1	0	0	0

Tabla 8.8: Tabla de frecuencias de MMDD con significado ilocutivo

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
		TOTAL	5	0	0	0
SA	Acceptation	faltaría más	1	0	0	0
		TOTAL	1	0	0	0
SA	Ask	de eso nada	2	0	0	0
SA	Ask	nanai	1	0	0	0
SA	Ask	naranjas de la China	1	0	0	0
SA	Ask	ni de coña	6	0	0	0
		TOTAL	10	0	0	0
SA	Assertion	como si	19	10	6	0
SA	Assertion	si	881	288	141	41
SA	Assertion	si acaso	1	3	3	1
		TOTAL	901	301	150	42
SA	Compliment	¡ole!	0	0	1	0
		TOTAL	0	0	1	0
SA	Assent	¡ajá!	8	0	1	0
		TOTAL	8	0	1	0
SA	Reciprocal Ac- tion	igualmente	1	1	2	0
		TOTAL	1	1	2	0
SA	Interaction Close	venga	3	0	1	0
		TOTAL	3	0	1	0
SA	Risk	qué demonios	0	0	1	0
		TOTAL	0	0	1	0
SA	Justification	el caso es que	9	0	3	0
SA	Justification	la cosa es que	2	0	0	0
SA	Justification	la cuestión es que	7	0	0	0
SA	Justification	lo que pasa es que	29	10	5	0
SA	Justification	lo que pasa que	32	4	1	1
SA	Justification	es que	395	86	26	2
		TOTAL	474	100	35	3
SA	Complaint	hay que ver	0	0	1	0
SA	Complaint	que si	31	10	4	0
		TOTAL	31	10	5	0
SA	Resignation	en fin	28	17	11	0
		TOTAL	28	17	11	0
SA	Greeting Open	buenos días	2	2	20	27
SA	Greeting Open	buenos días a todos	0	0	2	0
SA	Greeting Open	hola	20	2	15	2
SA	Greeting Open	muy buenos días	0	0	12	0
SA	Greeting Open	qué tal	25	0	0	2
		TOTAL	47	4	49	31
SA	Refuse	una mierda	2	0	0	0

Tabla 8.8: Tabla de frecuencias de MMDD con significado ilocutivo

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
		TOTAL	2	0	0	0
SA	Greeting Close	buenas noches	0	0	26	0
SA	Greeting Close	hasta luego	20	0	9	7
SA	Greeting Close	muy buenas noches	0	0	7	0
SA	Greeting Close	adiós	1	0	2	7
		TOTAL	21	0	44	14
SA	Gratitude	muchas gracias	6	14	8	12
SA	Gratitude	gracias	4	19	10	27
		TOTAL	10	33	18	39
SA	Action	¡anda!	19	2	2	0
SA	Action	¡hala!	17	0	0	0
SA	Action	hala	1	0	0	0
SA	Action	venga	33	3	6	0
SA	Action	adelante	8	4	3	0
SA	Action	anda	10	0	0	0
SA	Action	ya sabes	7	0	0	0
		TOTAL	95	9	11	0
SA	Warning	¡ojo!	0	0	4	0
SA	Warning	tú verás	1	0	0	0
SA	Warning	ya lo verás	2	0	0	0
SA	Warning	ya verás	6	1	0	0
SA	Warning	ya verás tú	2	0	0	0
SA	Warning	atención	5	6	3	0
SA	Warning	cuidado	4	3	1	0
SA	Warning	ojo	1	0	1	0
		TOTAL	21	10	9	0
SA	Apologize	lo siento mucho	2	0	0	0
SA	Apologize	perdona	10	5	3	0
SA	Apologize	perdóname	0	0	1	0
SA	Apologize	perdóname un momen- to	1	0	0	0
SA	Apologize	lo siento	2	1	0	0
SA	Apologize	perdón	10	11	5	0
		TOTAL	25	17	9	0
		TOTAL	1691	502	347	129

Tabla 8.8: Tabla de frecuencias de MMDD con significado ilocutivo

8.4.6. Frecuencia de marcadores de Deixis

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DEX	Contextual	a tomar por culo	3	0	0	0
DEX	Contextual	verá	2	0	0	0
		TOTAL	5	0	0	0
DEX	Social Interlocutor	¡anda!	19	2	2	0
DEX	Social Interlocutor	¡eh!	5	0	0	0
DEX	Social Interlocutor	¡ey!	4	0	0	0
DEX	Social Interlocutor	a que sí ?	8	0	0	0
DEX	Social Interlocutor	ah ?	9	0	0	0
DEX	Social Interlocutor	así que mira	2	0	0	0
DEX	Social Interlocutor	cómo ?	6	0	0	1
DEX	Social Interlocutor	cómo lo ves ?	1	0	0	0
DEX	Social Interlocutor	cómo que no ?	2	1	0	0
DEX	Social Interlocutor	de eso nada	2	0	0	0
DEX	Social Interlocutor	de verdad ?	1	0	0	0
DEX	Social Interlocutor	eh ?	219	154	56	4
DEX	Social Interlocutor	entiendes ?	1	0	0	0
DEX	Social Interlocutor	entonces ?	8	1	0	0
DEX	Social Interlocutor	eso mismo	1	0	0	0
DEX	Social Interlocutor	fíjate	26	1	11	0
DEX	Social Interlocutor	hija mía	2	0	1	0
DEX	Social Interlocutor	lo ves ?	3	0	0	0
DEX	Social Interlocutor	me entiendes ?	5	0	0	1
DEX	Social Interlocutor	me entiendes lo que te quiero decir ?	1	0	0	0
DEX	Social Interlocutor	no te acuerdas ?	5	0	0	0
DEX	Social Interlocutor	no te digo nada	2	0	0	0
DEX	Social Interlocutor	no ves ?	1	1	0	0
DEX	Social Interlocutor	no ?	417	167	181	6

Tabla 8.9: Tabla de frecuencias de MMDD de Deixis

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DEX	Social Interlocutor	o algo ?	3	0	0	0
DEX	Social Interlocutor	o cómo ?	3	0	0	0
DEX	Social Interlocutor	o no ?	27	2	3	0
DEX	Social Interlocutor	o qué ?	12	1	0	0
DEX	Social Interlocutor	o sí ?	0	0	1	0
DEX	Social Interlocutor	oiga	3	1	1	0
DEX	Social Interlocutor	os acordáis ?	4	0	0	0
DEX	Social Interlocutor	oye	89	7	24	2
DEX	Social Interlocutor	perdona	10	5	3	0
DEX	Social Interlocutor	perdona ?	2	0	0	0
DEX	Social Interlocutor	perdóname	0	0	1	0
DEX	Social Interlocutor	perdóname un momento	1	0	0	0
DEX	Social Interlocutor	por favor	14	4	3	582
DEX	Social Interlocutor	porfa	2	0	0	0
DEX	Social Interlocutor	qué creéis ?	1	0	0	0
DEX	Social Interlocutor	qué dices ?	3	0	0	0
DEX	Social Interlocutor	qué pasa ?	26	3	1	1
DEX	Social Interlocutor	qué te iba a decir ?	5	0	0	0
DEX	Social Interlocutor	queda claro ?	1	0	0	0
DEX	Social Interlocutor	queda claro todo esto ?	1	0	0	0
DEX	Social Interlocutor	sabes ?	167	2	0	1
DEX	Social Interlocutor	seguro ?	3	0	0	0
DEX	Social Interlocutor	sí ?	90	3	5	3
DEX	Social Interlocutor	sí o no ?	5	0	0	0
DEX	Social Interlocutor	te acuerdas ?	3	0	0	0
DEX	Social Interlocutor	vale ?	32	5	1	2

Tabla 8.9: Tabla de frecuencias de MMDD de Deixis

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DEX	Social Interlocutor	verás	6	0	0	0
DEX	Social Interlocutor	verdad hijo ?	1	0	0	0
DEX	Social Interlocutor	ves ?	27	0	1	0
DEX	Social Interlocutor	ves cómo es cierto ?	1	0	0	0
DEX	Social Interlocutor	y qué pasa ?	2	1	0	0
DEX	Social Interlocutor	y ya está ?	1	0	0	0
DEX	Social Interlocutor	ya lo verás	2	0	0	0
DEX	Social Interlocutor	ya verás	6	1	0	0
DEX	Social Interlocutor	ya verás tú	2	0	0	0
DEX	Social Interlocutor	ya ves	9	1	0	0
DEX	Social Interlocutor	ya ves que	1	0	0	0
DEX	Social Interlocutor	¡hombre!	13	0	0	0
DEX	Social Interlocutor	hombre	41	0	12	0
DEX	Social Interlocutor	venga	36	3	7	0
DEX	Social Interlocutor	a ver	67	17	5	1
DEX	Social Interlocutor	adelante	8	4	3	0
DEX	Social Interlocutor	anda	10	0	0	0
DEX	Social Interlocutor	atención	5	6	3	0
DEX	Social Interlocutor	cariño	4	0	2	0
DEX	Social Interlocutor	cuidado	4	3	1	0
DEX	Social Interlocutor	escucha	3	5	1	0
DEX	Social Interlocutor	hija	10	1	0	2
DEX	Social Interlocutor	macho	8	1	0	0
DEX	Social Interlocutor	maja	6	0	0	0
DEX	Social Interlocutor	majo	0	0	1	0
DEX	Social Interlocutor	mira	76	13	1	2

Tabla 8.9: Tabla de frecuencias de MMDD de Deixis

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DEX	Social Interlocutor	mujer	15	0	1	0
DEX	Social Interlocutor	ojo	1	0	1	0
DEX	Social Interlocutor	perdón	10	11	5	0
DEX	Social Interlocutor	señor	15	17	8	0
DEX	Social Interlocutor	tía	95	0	0	0
DEX	Social Interlocutor	tío	48	0	2	0
DEX	Social Interlocutor	vamos a ver	25	14	12	0
DEX	Social Interlocutor	verdad ?	6	4	2	0
DEX	Social Interlocutor	usted	17	16	22	0
DEX	Social Interlocutor	adiós	1	0	2	7
		TOTAL	1839	478	386	615
DEX	Textual Contextual	a este respecto	0	0	1	0
DEX	Textual Contextual	al respecto	1	1	2	0
DEX	Textual Contextual	ante todo	0	1	0	0
DEX	Textual Contextual	antes de nada	0	1	0	0
DEX	Textual Contextual	antes que nada	0	0	2	0
DEX	Textual Contextual	aparte de eso	2	0	1	0
DEX	Textual Contextual	así las cosas	0	0	1	0
DEX	Textual Contextual	así que por eso te digo	0	0	0	1
DEX	Textual Contextual	así y todo	1	0	0	0
DEX	Textual Contextual	aún así	5	1	0	0
DEX	Textual Contextual	de ahí que	0	1	0	0
DEX	Textual Contextual	si es que eso	2	0	0	0
DEX	Textual Contextual	anteriormente	5	7	0	0
DEX	Textual Contextual	aparte de	1	4	2	0
DEX	Textual Contextual	aparte	6	4	0	0
DEX	Textual Contextual	así es	1	3	1	0

Tabla 8.9: Tabla de frecuencias de MMDD de Deixis

Tipo	SubTipo	MD	INFO	FORM	Media	TELF
DEX	Textual Contextual	de ahí	12	7	2	0
DEX	Textual Contextual	de aquí	7	6	1	0
DEX	Textual Contextual	de entrada	2	2	2	0
DEX	Textual Contextual	de un lado	0	1	0	0
DEX	Textual Contextual	esto es	10	20	9	0
DEX	Textual Contextual	por ello	0	2	0	0
DEX	Textual Contextual	por eso	31	18	6	1
DEX	Textual Contextual	por otro	0	1	0	0
DEX	Textual Contextual	por un lado	2	12	3	0
DEX	Textual Contextual	primero	17	17	5	0
DEX	Textual Contextual	pues eso	35	4	2	0
DEX	Textual Contextual	segundo	6	4	3	0
DEX	Textual Contextual	tercero	0	4	0	0
DEX	Textual Contextual	y demás	4	9	1	0
DEX	Textual Contextual	y eso que	2	0	0	0
DEX	Textual Contextual	y eso	60	17	2	0
DEX	Textual Contextual	y nada	109	3	1	0
DEX	Textual Contextual	y todo	21	2	0	1
		TOTAL	342	152	47	3
		TOTAL	2186	630	433	618

Tabla 8.9: Tabla de frecuencias de MMDD de Deixis

Primeramente, la primera generalización que se puede obtener del análisis de las tablas es que hay una frecuencia de los fenómenos estudiados (y por tanto de uso de piezas léxicas en su favor) mucho más alta en la parte informal que en la formal del corpus.

En segundo lugar, como ya se ha advertido solo vamos a comentar brevemente algunas informaciones de cada fenómeno. Con cual, a partir de ahora, cada párrafo del análisis se corresponderá con un fenómeno pragmático y sus diferentes valores o subtipos.

Si comenzamos analizando los datos de frecuencia de los marcadores del discurso de las operaciones discursivas cabe destacar tres ideas. La primera es que hay operaciones que son eminentemente orales, como son las que se ocupan de finalizar el discurso (y *punto*) o de atenuar la transición de una secuencia temática a otra (indefinite transitional end, y *nada, pues eso, pues nada*). En segundo lugar, cabe mencionar que hay un predominio de los marcadores del discurso encargados de la ordenación textual en la parte formal frente a la parte informal. Esto confirmaría las reflexiones del J. Portolés en torno a la idea de que hay marcadores que necesitan un tratamiento mayor de la memoria a corto plazo, algo que solo se puede conseguir cuando se han interiorizado los mecanismos cognitivos procedentes de la escritura. El uso de los ordenadores implica una conciencia discursiva metaforizada en partes, mientras que los textos orales prefieren ir aportando la información recurriendo a recursos de unión de menor alcance textual. En tercer lugar, hay operaciones discursivas como la reformulación en donde se ve claramente una especialización de marcadores que se prefieren para lo oral, como es el caso de *o sea* que presenta 562 ocurrencias en la parte informal y 142 en la parte formal, o los marcadores *es decir, digamos, esto es* cuyas frecuencias lo sitúan más dentro de lo formal. Podemos hacer lo mismo para todas los valores, pero creemos que el caso de la reformulación puede ser un ejemplo suficiente para ilustrar cómo se pueden leer las tablas. En último lugar, destacaremos comportamientos interesantes de los marcadores de consecuencia, en los que los conectores *pues* y *entonces* parecen ser los elegidos por la modalidad oral para expresar la consecuencia, con una diferencia de uso de 2174 frente a 1063 ocurrencias en la parte formal (más influida por la modalidad escrita), que prefiere otros marcadores como *por lo cual, con lo cual, por lo tanto* o *por tanto*.

En lo que respecta a las palabras que codifican evaluaciones emocionales positivas, negativas o de sorpresa, lo primero que debemos decir es que estos tres valores se dan con más frecuencia en la parte informal que en la formal, y que dentro de cada uno de ellos, las evaluaciones negativas tienen en la parte formal una tasa más elevada que la expresión de evaluaciones positivas y la sorpresa, la cual es eminentemente informal. Estos datos pueden ser muy interesantes en los futuros estudios contrastivos con otras lenguas, en lo que a la verbalización de emociones a través de estas palabras se refieren, y pueden contribuir a definir quizás más concretamente la famosa cortesía positiva de la cultura española frente

a otras culturas.

Siguiendo con las cuestiones sociales, pasemos a estudiar los resultados de las operaciones de intensificación, atenuación e interacción. Aquí también como hemos señalado hay un mayor número de ocurrencias en los textos no planificados que en los formales. Pero dentro de esta generalización, como en el caso de las operaciones discursivas, hay perfiles de marcadores del discurso que son netamente orales, y con altísimas frecuencias de uso, como es el caso de *bueno, claro, vale, sabes?, sí?, no? venga, y nada, etc.*; mientras que hay marcadores que como *obviamente, lógicamente, en absoluto, indudablemente o naturalmente* que tienen una presencia mayor en el género formal. Aparte de esto, destaquemos la frecuencia de uso prácticamente igualada en las dos partes en el caso de marcadores como *de hecho*, que en principio, haciendo caso de mi competencia lingüística, hubiera apostado por una mayor frecuencia en el género formal. Lo mismo ocurre con marcadores como *eh?, en absoluto, por supuesto y, en menor medida, realmente*.

Pasemos de largo sobre los valores de evidencialidad y actos de habla, ya que es una tipología que deberá ampliarse en el futuro, ya que los datos de frecuencia no son relevantes respecto de nada, ya que también hay actos de habla en los enunciados que no hemos marcado, y marcas de evidencialidad sintácticas que no están reflejadas en la anotación, con lo que los datos en estos valores de momento no presentan, a efectos de lo breve del comentario, una relevancia mayor.

Terminemos este apartado con las cifras de frecuencia de dos marcadores que en el etiquetado, pese a su posible carácter polémico, han sido etiquetados como deícticos, y que guardan bastante semejanza de significado; dichos marcadores son *a ver* y *vamos a ver*; ya en otras ocasiones se ha apuntado que estos marcadores presentan diferencias de registro. En las tablas del corpus podemos consultar que el marcador *a ver* aparece 68 veces en la parte informal del corpus, frente a 22 veces en la parte formal; mientras que en el caso de *vamos a ver* encontramos una frecuencia menor y más igualada en los dos partes: 25 veces en informal y 26 veces en la parte formal. Sirva este pequeño apunte que se extenderá en una publicación en el futuro, para reforzar la idea de que es bastante posible que la diferencia entre uno y otro marcador sea fundamentalmente sociolingüística.

Una vez terminado este breve análisis de los resultados, que no hemos hecho más amplio por razones de espacio, pasemos a cerrar el capítulo.

8.5. Conclusiones

Permítanme, antes de pasar a la segunda aplicación, resumir algunas cuestiones interesantes de este capítulo. En primer lugar, me gustaría volver a subrayar la necesidad de apoyarse en el conocimiento de las técnicas de procesamiento computacional de información para conferir autoridad a los estudios lingüísticos de corte cuantitativo. Debería formar parte de las exigencias de empiricidad en los comités de evaluación explicitar toda la información referente al modo en que se ha gestionado la información de los corpus, hecho el etiquetado de la información y la cuantificación posterior de los datos. Las innovaciones en el terreno de lo computacional deben ayudarnos a hacer Lingüística.

En segundo lugar, me gustaría haber contribuido con este estudio a un área bastante inexplorada en los estudios del español oral como es el análisis de los marcadores discursivos en corpus desde una perspectiva cuantitativa. Al tiempo que espero haber ilustrado con suficiente claridad cómo la anotación de corpus nos abre un abanico de numerosas posibilidades para el lingüista que decida estudiar un fenómeno determinado apoyándose en datos de la lengua reales, aparte de en su propia competencia lingüística.

En tercer lugar, es de rigor hacer un último comentario con respecto al conocimiento que se puede construir a partir de los datos extraídos. Las tablas de frecuencia efectivamente confirman una idea que los pragmlingüistas habían ya señalado, y es que los marcadores discursivos también están sometidos a las exigencias contextuales de registro comunicativo. Aquí hemos utilizado la variable de género discursivo, pero añadiendo pequeñas modificaciones en el programa utilizado, también podríamos estudiar la relación de frecuencia de los marcadores discursivos con otras variables contextuales especificadas en las cabeceras de los textos como por ejemplo la edad, el grado de educación formal de los hablantes o el género. Y así lo mostraremos en trabajos futuros.

En cuarto y penúltimo lugar, en lo que respecta al análisis de los resultados, no hemos presentado aquí conclusiones que se aparten sustancialmente de lo que venía registrando de una metodología u otra la bibliografía en relación a la variante escrita y oral. Destáquese la especialización de marcadores en un género u otro y la mayor presencia de determinados fenómenos en una variante con respecto a la otra.

Por último, y atendiendo al hecho de que parte de la tesis pueda interesarle a los lectores con una formación no lingüística, sirva este capítulo para ejemplificar esta visión de la lenguas en las que se viene insistiendo a lo largo de la tesis como una masa de posibilidades conformada por grandes tendencias y absolutamente susceptible a las variaciones del entorno.

Pasemos pues al penúltimo capítulo de la tesis, donde se presenta una aplicación del

corpus anotado con este tipo de información a la Enseñanza de español a extranjeros.

Capítulo 9

La enseñanza por ordenador de los marcadores del discurso a alumnos italianos con C-ORAL-ROM, un corpus del español oral

Sólo son educadas las personas que han aprendido cómo aprender (NEBRIJA, 1998)

9.1. Introducción

En más de una ocasión la Unión Europea ha manifestado su preocupación por la creación de una sociedad multilingüe (MARÍN, 2002). Es posible que esa fuera, entre otras, una de las razones por las que financió la creación de estos corpus de habla oral con el proyecto C-ORAL-ROM. No obstante, las muestras recogidas en los corpus deben someterse primero a un tratamiento didáctico antes de que puedan resultar útiles a los estudiantes de otras partes de la Unión Europea que se puedan servir de estos datos para mejorar su conocimiento de otras lenguas europeas.

Otro hecho que evidencia esta preocupación de la política europea por la Europa Multilingüe ha sido la elaboración de un estándar de evaluación, el **Marco europeo ELE**, que intenta especificar con detalle qué tipo de conocimientos y habilidades conforman cada uno de los niveles que se han marcado: A1 A2 B1 B2 C1 C2.

El marco europeo ELE adopta un punto de vista funcional de la lengua y, consecuentemente, se centra en las “funciones comunicativas que permiten la interacción en situaciones de la vida real” (NEBRIJA, 1998). Es ya un lugar común en esta disciplina señalar el abandono en los manuales de la gramática y la apuesta por los enfoques comunicativos, esto es, aquellos que privilegian el uso de la lengua en contexto. Dicha apuesta hace que todos

los conocimientos procedentes de la Pragmática se vuelvan muy atractivos para el profesor de ELE. Pero, por contrapartida, esta disciplina hereda las carencias de una Pragmática de reciente creación, muy fuerte en su base teórica, muy rica en la descripción, pero muy poco preocupada por crear modelos unitarios y estándares susceptibles de ser aplicados a otras disciplinas como la que vamos a tratar aquí.

En este capítulo desarrollamos una de las posibles potencialidades de un corpus oral del español anotado con información pragmática. En esta aplicación intentamos conciliar tres ramas de actualidad en este campo: la enseñanza por ordenador, la enseñanza de español como lengua extranjera, y por último, la Lingüística de corpus. El resultado de esta intersección de disciplinas ha sido el diseño de una interfaz en la Web que sirva de intermediario entre el estudiante de español y la información del corpus. Dicha interfaz puede consultarse en el CD que acompaña a este trabajo; lo que encontrará el lector en este capítulo será un breve estado de la cuestión en materia de marcos teóricos que han predominado en este campo de estudio y la concepción teórica que subyace al diseño del interfaz.

Antes de comenzar, es de rigor mencionar que este trabajo se enmarca dentro de una de las líneas de investigación abiertas en este momento en el Laboratorio de Lingüística Informática de la UAM. Dentro de esta línea, se diferencia por un lado la creación de un Learner Corpora, y, por otro, la adaptación del corpus C-ORAL-ROM a la enseñanza de español como segunda lengua, proyecto denominado C-ORAL-ROM ELE (CAMPILLOS y otros, 2007) (GONZÁLEZ-LEDESMA, 2007b).

9.2. Estado de la cuestión: ELE, Pragmática, ordenadores y corpus

9.2.1. Marcos teóricos y metodológicos para ELE

La historia metodológica en la enseñanza de lenguas extranjeras ha pasado por tres fases: el enfoque conductista, el enfoque comunicativo y el integrado. Este último modelo impera desde los años noventa y se basa en una visión de tipo más social o socio-cognitiva del aprendizaje, fundamentado en la psicología de Vigotsky de 1964, el cual relaciona el desarrollo cognitivo del individuo con las estructuras sociales del aprendizaje en colaboración.

A continuación, vamos a destacar las principales aportaciones teóricas de los diversos marcos que han dominado a lo largo de la historia el panorama de esta disciplina, y sobre cuyos principios hemos intentado construir nuestro trabajo:

9.2.1.1. De las vertientes cognitivo-sociales de corte constructivista

La teoría cognitiva nace en el campo de la Psicología Cognitiva y se centra en el estudio del papel de los procesos cognitivos más generales involucrados en la adquisición de la lengua. A continuación, enumeramos las ideas más interesantes que ha aportado esta línea al conocimiento del aprendizaje de lenguas (NEBRIJA, 1997) y (BEAU-FLY, 1995) :

1. Importancia del papel de la reflexión metalingüística y de la metacognición en el aprendizaje de lenguas.
2. Conciencia del valor social del uso lingüístico.
3. Constructivismo: los conocimientos se aprenderán a medida que el desarrollo de las capacidades lo vayan permitiendo. La clave del éxito es conectar/activar el conocimiento previo. El aprendizaje se da en etapas pero es recursivo.
4. Incidencia en las dimensiones del pensamiento: comprensión, composición, resolución de problemas, decisiones, pensamiento crítico y creativo.
5. Prototipo del buen alumno: aquel que tiene un rol activo en el aprendizaje. El alumno aprende más de una actividad si está regida, dirigida por él mismo. Por tanto:
 - a) El alumno debe regular su aprendizaje.
 - b) Debe ser capaz de tener acceso a lo que sabe.
 - c) Es obligada la reflexión sobre los estados emocionales del alumno en términos de confianza y seguridad.
 - d) La lengua no se aprende por imitación de modelos sino por la presentación de hipótesis.
 - e) El profesor deja de ser el centro.
6. El error no recibe un tratamiento negativo.

9.2.1.2. Del enfoque comunicativo

De este enfoque, estas son las aportaciones más interesantes en lo que respecta a la mejor manera de aprender una lengua extranjera (NEBRIJA, 1998), (MONTOLÍO, 2004) e (NEBRIJA, 1997):

1. Foco de atención en el uso.
2. Mayor importancia al contexto.
3. Una visión funcional de las formas lingüísticas. Ejemplos de esta visión funcional sería el uso el imperfecto de subjuntivo para expresar deseo o la co-ocurrencia de los marcadores como *pero si* para expresar desacuerdo.
4. La lengua es un instrumento de comunicación: es un instrumento útil para la actuación y la resolución de problemas en la vida cotidiana.
5. La comunicación es un proceso de intercambio de información y de negociación de significado. Nosotros añadimos que el intercambio de información es una excusa para intercambiar afectos, emociones y crear lazos de solidaridad social. Por ejemplo, poco importa de qué hablas cuando uno está solo y quiere hacerse amigos en un país extranjero, lo importante es que se está hablando.
6. La enseñanza es un proceso.
7. Visión del alumno: el estudiante es responsable de su aprendizaje.
8. Es importante que el estudiante perciba que con el uso de la lengua va a lograr algo. Hay que crear necesidades comunicativas que luego reproducirán en el futuro.

9.2.1.3. Del enfoque por tareas

Este enfoque consiste en proporcionar un *input* al alumnado que le servirá no tanto como de modelo, sino como una muestra de lengua para analizar cómo se expresa una idea o sentimiento. A partir de este análisis y reflexión, el alumno será capaz de diseñar un productor final (*outcome*) como resultado de todo su proceso de aprendizaje (COTS~i CAMINOS y otros, 2007). La tarea se centra más en el significado que en la forma lingüística. Te preparan para saber qué decir en situaciones reales.

En cuanto a la definición de tarea, entendemos por esta última cualquier iniciativa para el aprendizaje que consista en la realización en el aula de actividades de uso de la lengua representativas de las actividades que se llevan a cabo fuera de del aula en la vida real. La tarea posee determinadas características:

- Estructura pedagógica
- Estructura abierta a la intervención de los alumnos durante la ejecución de la tarea.
- Debe ser unitaria

- Toques de atención sobre la forma lingüística.
- Deber resultar próxima al alumnado desde un punto de vista del contenido representado.
- Debe estar vinculada a los objetivos del currículum.
- Debe ser evaluable.

Pasemos ahora a hablar de la relación entre Enseñanza de segundas lenguas, Pragmática y marcadores del discurso.

9.2.2. La enseñanza de la Pragmática y de los marcadores del discurso

¿Qué es la Pragmática? ¿Es el estudio del significado desde diferentes disciplinas? ¿Es una dimensión de estudio o un módulo de la gramática? ¿Es aquello que estudia la relación entre lo que el hablante quiere decir y lo que el receptor entiende? ¿Es lo que refleja las relaciones sociales? ¿Qué Pragmática enseñamos, la social o la cognitiva? ¿Debemos enseñar formas lingüísticas o más bien estrategias? Si como dice Verschueren, el lenguaje es una variable negociable, ¿cómo hacemos para que el alumno entienda que el lenguaje es un instrumento y no una enciclopedia? Ya se apuntó en el capítulo dos que la falta de consenso con respecto a qué lugar ocupa la Pragmática en la lengua dificulta después la transmisión de este conocimiento no solo a los profesores de ELE, muchos de los cuales no poseen una formación fuerte en Lingüística, y menos en Pragmática, sino también a los alumnos.

No es necesario recurrir a las fuentes de conocimiento escritas que han tratado el tema (HAVERKATE, 1998) y (GALLARDO, 1997) para saber que un error de tipo gramatical se perdona y, sin embargo, uno pragmático, no. Pero aun así, citemos las palabras de P. BOU-FRANCH al respecto:

“Los alumnos deben por lo menos estar sensibilizados al respecto y ser conscientes de que ciertos comportamientos pueden ser malinterpretados y afectar de manera directa a la imagen de sí mismos que proyectan y cómo van a ser evaluados por sus interlocutores.”(BOU-FRANCH y GARCÉS, 1994)

La codificación opaca de normas sociales y culturales en las formas lingüísticas hace que la enseñanza de lo que Gallardo Paúls (GALLARDO, 1997) denomina **el componente**

pragmático sea una tarea bastante difícil siquiera de afrontar. Como es de esperar, hay un acuerdo teórico en las fuentes consultadas sobre la importancia de esta dimensión a la hora de enseñar las segundas lenguas, pero, por contrapartida, y solo si se nos permite la expresión, brillan por su ausencia los marcos metodológicos para enseñar estos conocimientos de una forma integrada, como parte del sistema lingüístico. Tan sólo hace unos pocos años el componente cultural ha empezado a hacer acto de presencia en los manuales ELE pero siempre de una manera separada y muy poco o nada asociada a la forma lingüística (NEBRIJA, 1998). Como decíamos en la introducción, la no sistematización de lo que Lotman denomina la socioesfera, esto es, el conjunto de normas sociales y culturales, impide la necesaria integración del conocimiento lingüístico, comunicativo, sociocultural y estratégico, y que se ha denominado **competencia pragmática y/o comunicativa** ¹.

9.2.2.1. Enseñanza de marcadores del discurso en ELE

Los marcadores del discurso son vistos como indicadores del nivel de dominio de la lengua del hablante no nativo (CASTILLO, 2003) y (PENADÉS, 1997). Al mismo tiempo, los autores que se han ocupado de esta intersección disciplinar reconocen que la enseñanza de estas partículas constituye una de las carencias más graves que sufre actualmente la enseñanza de segundas lenguas (LAMIROY, 1994). Entre las dificultades para enseñarlas, Penades señala el problema de la equivalencia comunicativa y la incapacidad para explicar sistemáticamente las unidades fraseológicas, muchas de las cuales, como sabemos son marcadores del discurso. Por su parte M. Zorraquino (MARTÍN y MONTOLÍO, 1998) subraya “la necesidad de aplicar los conocimientos que vamos obteniendo sobre estas unidades, tan importantes desde el punto de vista comunicativo, a la Enseñanza del español como lengua extranjera”. Citemos algunas investigaciones sobre la comprensión y uso de los marcadores por parte de los hablantes de una segunda lengua.

J. M. FULLER (LAHUERTA y PELAYO, 2003), demostró en sus investigaciones cómo el desconocimiento del significado de los marcadores del discurso afecta a la comprensión durante la lectura de textos. En (FULLER, 2003) se compara el uso de marcadores *oh*, *well* y *you know?* en hablantes nativos y no nativos en dos géneros discursivos: las entrevistas y las academias. En el estudio, se concluyó que los hablantes nativos hacen un uso mayor de los marcadores *oh* y *well* en las conversaciones mientras que aumenta el uso de *you know* en las entrevistas. Mientras que por su parte los hablantes no nativos tienen una tasa de

¹**Definición del concepto de competencia comunicativa** según Hymes: capacidad de usar la lengua de forma adecuada a cada situación concreta de la comunicación. El concepto de competencia ha sido entendido por Canale y Swain como una macrocompetencia que a su vez se divide en subcompetencias: competencia discursiva, gramatical, sociológica y estratégica. (NEBRIJA, 1998)

frecuencia baja y formulaica de marcadores discursivos e independiente del género discursivos (MARCHANTE, 2005).

Además también se ha investigado la interferencias que en el aprendizaje de marcadores del discurso tienen los marcadores de la lengua nativa. Los españoles no usan *on the other hand* con un significado contrastivo sino más bien con el significado de *moreover* o *furhtermore* o *anyway* (MURILLO, 1998).

Finalmente, parece que hay un acuerdo en afirmar que enseñar marcadores implica un enfoque comunicativo pero “no hay acuerdo sobre los marcadores, y sus significado ni sobre qué perspectiva adoptar si teórica o aplicada.”

“El profesor ELE no encontrara en los distintos manuales publicados sobre la enseñanza del español, algunos de ellos, los mas actuales, se basan en los contenidos según el Marco Común Europeo de Referencia de las lenguas, A1 (acceso) y A2 (plataforma), ninguna explicación ni introducción sobre estas unidades”(MARCHANTE, 2005).

Respecto de la clasificación de estos marcadores, Marchante se queja de que “muchos manuales de ELE presentan tanto a profesores como estudiantes muy diferentes clasificaciones con respecto a los marcadores argumentativos [...] en dichas clasificaciones no presentaron muchos casos las verdaderas diferencias que entre ellos existe”(MARCHANTE, 2005). Hay un dilema bastante difícil de resolver en torno a qué clasificación de marcadores presentar, y sobre hasta qué punto se deben enseñar las diferencias entre marcadores de la misma clase. Coincidimos con S. Murillo (MURILLO, 1998) en que hay que tomar clasificaciones de marcadores discursivos que valgan para más de un campo de conocimientos, no solo para este área sino también para la Psicolingüística o la comunicación hombre-máquina.

Así pues, para resumir, hay un acuerdo general en la necesidad de enseñar la dimensión pragmática de la lengua en general, y el uso de estas partículas discursivas en particular, aunque faltan modelos y propuestas didácticas efectivas que superen esta enorme carencia en la enseñanza de segundas lenguas a las que asistimos por el momento en esta disciplina.

9.2.3. La enseñanza de lenguas y el ordenador

A continuación enumeraremos las ventajas y desventajas que se han señalado respecto de la enseñanza de lenguas por ordenador ((BERMEJO y CASADO, 2000) y (RODRÍGUEZ~MARTÍN, 2004)):

1. La ventajas de aprender con el ordenador:

- a) El ordenador facilita la concentración que la clase durante más tiempo.
- b) Se reduce la inhibición a hablar y el miedo al ridículo.
- c) Se considera positivo para el aprendizaje la presencia de estímulos sonoros y virtuales como, por ejemplo, los colores. Según (BERMEJO y CASADO, 2000) hemos pasado del *homo sapiens* al *homo videns*, dominado por la cultura de la imagen y de lo virtual.
- d) Retroalimentación.
- e) Aprender haciendo reduce el riesgo de no recordar lo aprendido.

2. Las críticas que ha recibido la Enseñanza de lenguas por ordenador:

- a) El alumno puede perderse navegando sin rumbo fijo.
- b) La falta de base pedagógica de los productos comerciales, y cuando la hay, es una metodología tradicional, fenómeno que ha recibido el nombre de **ciber estructuralismo**. Hay que meditar los contenidos que se informatizan, y adaptarlos al medio de comunicación y a los nuevos marcos teóricos de ELE.
- c) La tecnologías multimedia presentan limitaciones en la valoración y corrección de las producciones del alumno.

9.2.4. La enseñanza de segundas lenguas con corpus

Los materiales didácticos muchas veces son el único *input* de datos con los que cuenta el alumno de L2(GONZÁLEZ, 2004). Un corpus se puede enfocar también como material didáctico para la enseñanza de lenguas, y además tiene el atractivo de ser muestras directas, reales de la lengua.

Carlota Nicolás (CURTO y otros, 2007) enseña español a alumnos italianos en la universidad de Florencia. Esta profesora tiene una larga experiencia como docente, y lleva años también investigando los beneficios para el aprendizaje de la enseñanza de segunda lenguas con corpus. Escuchemos algunas de sus reflexiones en las que defiende los corpus como parte integrante del material didáctico del alumno (CURTO y otros, 2007):

“Me pregunto por qué se considera esta lengua *teatral* [a propósito de las muestras de lengua de los manuales] más oportuna o adecuada de enseñar que la lengua real; quizás se

cree que la complejidad de esta última es mayor, pero si como profesores de español estudiamos los corpus del oral apreciaremos las diferencias y muy probablemente lleguemos a afirmar que en el uso de la lengua oral real el alumno puede encontrarse más relajado, pues no debe hacer que cada palabra que pronuncie sea informativa y tenga una gramaticalidad (características propias de la lengua escrita, especialmente de los textos orales de los manuales). El alumno que estudie la lengua oral llegará a saber que puede hacer frases nominales perfectamente informativas, puede perder el hilo, volver a empezar, cambiar programa informativo, puede repetir, ser redundante, hacer anacolutos, etc. El alumno tiene además que ser consciente, y esto es muy importante, que en la recepción de la lengua oral debe prestar mayor atención a la prosodia y tiene, además, que saber esperar para captar lo que es esencial en el discurso, para que este pueda proseguir en la interacción, para que se mantenga el discurrir del discurso sin impedir la comunicación.”(CURTO y otros, 2007)

Es digno de mención señalar el recurso electrónico accesible en internet *Diccionario de partículas discursivas en español*, proyecto que tiene como investigador principal a Antonio Briz, y donde podemos acceder a una rica explicación lingüística de un glosario limitado de marcadores del discurso con ejemplos (con sonido) procedentes diferentes corpus del español, entre ellos el CREA y los corpus del grupo VALESCO (BRIZ, 2008)

Una vez establecidos las bases de conocimiento, pasemos a explicar nuestro trabajo.

9.3. INTER ALIA, un interfaz Web para el aprendizaje de la pragmática

9.3.1. Declaración de principios

Primero de todo, me gustaría destacar las ideas principales en torno a las cuales va a girar la concepción de la aplicación informática que hemos desarrollado, y que pasaremos a enumerar a continuación y a desarrollar más detenidamente después:

1. Pragmática en el aula desde el primer día.
2. La lengua materna es un recurso metalingüístico.
3. Consciencia de la imagen pública.
4. Crítica al método comunicativo.

9.3.1.1. Pragmática y marcadores del discurso en el aula desde el primer día

En primer lugar, el conocimiento pragmático sobre una lengua debe comenzar a aprenderse desde el primer día en que una persona decide aprender dicha lengua. Tal y como hemos dicho en la primera parte del capítulo, las lenguas son un vehículo de comunicación social, y codifican partes que se refieren a ella, por ejemplo, la cortesía. Detrás de esto, hay un nivel cognitivo, uno se comunica para intercambiar su visión del mundo con los demás con el fin de cumplir sus metas, las cuales pueden ser de muy diversa naturaleza. En el caso de un estudiante, lo primero que debe aprender cuando llega a un lugar nuevo es a establecer una relación social exitosa que le permita una interacción comunicativa continuada en el tiempo. Si bien es difícil conocer las normas sociales que rigen una sociedad en la que no se ha vivido, también es cierto que podemos salvar estas distancias si aprendemos a utilizar aquellas palabras que una lengua tiene reservada para expresar las emociones, el acuerdo, el desacuerdo, para argumentar tu punto de vista, y en definitiva todas aquellas expresiones lingüísticas de las que nos servimos para establecer una relación con el otro, tales como por ejemplo, en español el marcador discursivo *o sea*, la interjección *la madre que te parió* o la evaluación positiva de la realidad con la locución *de puta madre*. Los marcadores del discurso se empiezan a enseñar en los niveles B1(umbral) B2 (avanzado) C1(dominio operativo) C2(maestría)(MARCHANTE, 2005), con lo cual el aprendizaje de estas partículas normalmente se retrasa a los últimas etapas del aprendizaje, una vez que se considera que el alumno tiene un conocimiento general del resto del vocabulario que en teoría se consideraba básico, como por ejemplo la palabra *mesa* (si bien es cierto, algunas de ellas, como el saludo, se enseña en los primeros niveles). No estoy de acuerdo con L. Gurillo en que la unidades fraseológicas (buena parte de los marcadores del discurso lo son) deban enseñarse en último curso. Estas palabras, que se aprenden muy tarde, son las que más se echan de menos cuando uno intenta ser uno mismo. El alumno, a falta del conocimiento de partículas que expresen emociones y operaciones complejas de pensamiento, de forma automática recurre a las interjecciones y a los conectores de su lengua madre, palabras que en un contexto extranjero solo él entiende. Utilizando el marco teórico de Penélope Brown, podría decirse que el beneficio en términos sociales el uso de los marcadores del discurso no tiene parangón respecto al desconocimiento, por ejemplo, de una conjugación verbal de subjuntivo o de una palabra determinada. Se hace más amigos con una interjección de evaluación positiva bien traída (*ole* y *ole*) que con una gramática sin errores.

9.3.1.2. La lengua materna es un recurso metalingüístico

A pesar de que haya sido una idea muy criticada, la lengua madre debe considerarse siempre un recurso, y no un conocimiento que se debe inhibir. La lengua madre puede ayudarnos a asociar los contextos a las expresiones. Así por ejemplo, en italiano cuando alguien te pide permiso para entrar a un sitio en el que estás tú, se dice *avanti*, mientras que en español, *pasa*. La lengua nativa del alumno debe estar presente en el diálogo mental sobre las diversas posibilidades de codificar el mundo.

Por tanto, propongo ir del fenómeno a las piezas léxicas. Esto es, primero explicar la operación de la contraargumentación y después ofrecer un conjunto de partículas que se encargan de llevar a cabo esta operación, tanto en español como en italiano, que es el perfil del alumno para el que está diseñado esta Web.

9.3.1.3. Consciencia de la imagen pública

Esta unidad está concebida para salvar los problemas de imagen pública que introducimos en la primera parte teórica, y sobre los cuales me gustaría insistir aquí un poco más. Como sabemos, en el transcurso de una conversación el individuo ofrece una impresión de su personalidad a sus interlocutores que puede ser puesta en tela de juicio (BOU-FRANCH y GARCÉS, 1994). No se suele mencionar en los manuales el trasfondo psicológico del aprendizaje de idiomas, cuya preocupación más importante es la presentación de nuestra persona en la vida cotidiana:

“Hay que tener en cuenta que a los hablantes no nativos especialmente a los jóvenes con bajo conocimiento de la lengua se les confiere un estatus similar al de los niños en la intervención a nivel pragmático” (BOU-FRANCH y GARCÉS, 1994).

No estoy de acuerdo con esta afirmación. Al bebé se le estimula desde un punto de vista emocional, se le quiere y se le aplaude cada vez que dice cualquier sonido que se pueda asemejar al código lingüístico del adulto. Con el adulto que aprende un idioma nuevo el tratamiento emocional es el inverso, con las consecuencias que tiene en la autoestima y en la presentación del individuo en sociedad. Al adulto se le presupone sentido común (algo eminentemente pragmático) mientras que al bebé, no. Además, ¿cuántas veces repetimos la palabra *bueno* a lo largo de nuestra vida? Un extranjero recién llegado la habrá pronunciado menos de diez veces quizás. Así que lo que para él es nuevo para el nativo es absolutamente evidente. Imaginemos qué ocurre cuando trascendemos el nivel de palabra y nos centramos

en los significados de los enunciados, como por ejemplo, un acto de habla o una estrategia cortés. El niño tiene dos años para hablar, y es un aprendizaje que lejos de ser criticado, ironizado o parodiado (como suele pasar en el caso del adulto) es por el contrario estimulado afectivamente. Estas cosas no se contemplan pero son importantes a la hora de aprender. Los estudiantes del árabe pasan por las mismas etapas cuando aprenden el sistema de escritura. De hecho suelen decir cosas del tipo *ya escribo como escribe en árabe un chaval de quince años*. Cuando somos adultos tenemos una imagen pública que mantener. Hace un tiempo una amiga italiana estudiante de francés me contó cómo en sus inicios en Lyon el primer amigo que hizo, después de tres meses sin entender prácticamente nada, le invitó a una fiesta. Este amigo, consciente de las limitaciones de mi amiga, le aconsejó que siempre que cuando la aludieran a ella directamente, preguntándole qué pensaba del asunto, que ella siempre contestará: *c'e pas faux* (No es falso). Esta estrategia de cortesía social está completamente relacionada con el compromiso del hablante con la verdad de los enunciados de la interacción. Un *no es falso* es una estrategia cortés de no asumir ninguna responsabilidad sobre lo dicho y por tanto, evita no tener que argumentar, algo valiosísimo para las primeras etapas, en las que lo importante es participar, nunca ser el centro de atención.

Así pues, esperamos haber dejado claro que la personalidad de uno se transforma en función de sus habilidades expresivas (ALONSO, 1993) y que aprender es un proceso emocional e interpersonal además de cognitivo.

9.3.1.4. Crítica al método comunicativo

¿Cómo aprender la Pragmática? Si bien estamos de acuerdo en que los contenidos deben presentarse de una manera funcional (por ejemplo, enseñar la intensificación, bajo el lema de aprende a reforzar tus afirmaciones), no estamos de acuerdo con el verdadero caos que existe en la enseñanza de español para extranjeros a la hora de enseñar lo que ellos han llamado las funciones comunicativas, y que lejos de constituir una clase cerrada, no presentan un estudio sistemático entre dicha función y las diferentes formas lingüísticas que lo llevan a cabo. Citemos las siguientes palabras de L. Cortés al respecto:

Al ser los objetivos comunicativos casi infinitos, casi infinitas serán también las posibles acciones verbales entre las cuales el hablante habrá de decantarse para organizar su discurso, que a su vez, está vinculado a la situación. Ello originará una enorme complejidad tipológica que, por un lado, lleva a la desesperación a los lingüistas docentes preocupados por estas cuestiones y, por otro, dificulta aun más la respuesta a preguntas tales como qué actividades de

observación, de manipulación, de producción y de evaluación deben llevarse a cabo para una enseñanza más eficaz de la lengua oral, por ejemplo. (CORTÉS, 2002)

Así, en la gramática de Matte Bonn (MATTE, 1992) se enseñan cosas del tipo *aprende a dar tu opinión*, enseñando expresiones como por ejemplo *en mi opinión*, *desde mi punto de vista*, pero no explican el fenómeno que está detrás, que es la señalización de las fuentes de información y la consecuente influencia o valor argumentativo que tiene a la hora de ganar autoridad en el discurso. A continuación, señalamos una lista de funciones comunicativas para que el lector se haga una idea de a qué nos estamos refiriendo y para reforzar aún más nuestro profundo desacuerdo con la falta de sistematicidad de los enfoques comunicativos en este sentido: expresar proporciones, expresar necesidad, referencias en grupo, conocimientos, ideas, disponibilidad de información, expresar esperanza, desear cosas buenas a los otros, antes de saber, después de saber, etc.

Nuestra interfaz está concebida para el nivel A. En este primer nivel se estudiarán los marcadores de acuerdo con los fenómenos a los que remiten pero no se explicarán las diferencias que hay entre los marcadores con una misma función.

Pasemos ahora a explicar el interfaz. Básicamente el interfaz se divide en dos módulos. En el primer módulo, el alumno podrá acceder a una breve definición de los fenómenos de carácter pragmático explicados en esta tesis, y a la lista de formas lingüísticas a través de las cuales dicho conocimiento se expresa. Cada una de estas formas, remiten a través de un link a un ejemplo de uso procedente del corpus. A su vez también puede acceder al modo en que dicho conocimiento pragmático se expresa en su lengua, para así tener una ayuda añadida a la hora de poder deducir el sentido o el uso de un marcador pragmático. La perspectiva contrastiva se introduce cuando ponemos marcadores que en la lengua nativa cumplen la misma función. Esto último no lo hacemos desde la perspectiva de la equivalencia semántica, sino como materializaciones lingüísticas de fenómenos semánticos de orden superior de carácter cognitivo-social. Además, el usuario también podrá consultar el lexicón de marcadores del discurso, también incluido en la página web, como ya señalamos en el capítulo cinco.

9.3.2. El interfaz Web: Inter Alia

Una web está sustentada en el concepto de hipertexto a la hora de mostrar la información. El hipertexto activa la iniciativa del alumno en el aprendizaje en función de sus inquietudes o necesidades comunicativas. Ya señalamos en el apartado teórico la importancia de la agentividad del alumno en el proceso de aprendizaje. Lo que el alumno va a encontrar en el interfaz será un sistema que le ayude a gestionar la información de un corpus según sus objetivos.

En nuestra propuesta, vamos desde el fenómeno a la forma lingüística. Todas las fichas tendrán la misma estructura:

1. Definición.
2. Lista de marcadores: en italiano y en en español.
3. Tasa de frecuencia en función de los géneros del discurso.
4. Consulta de corpus.

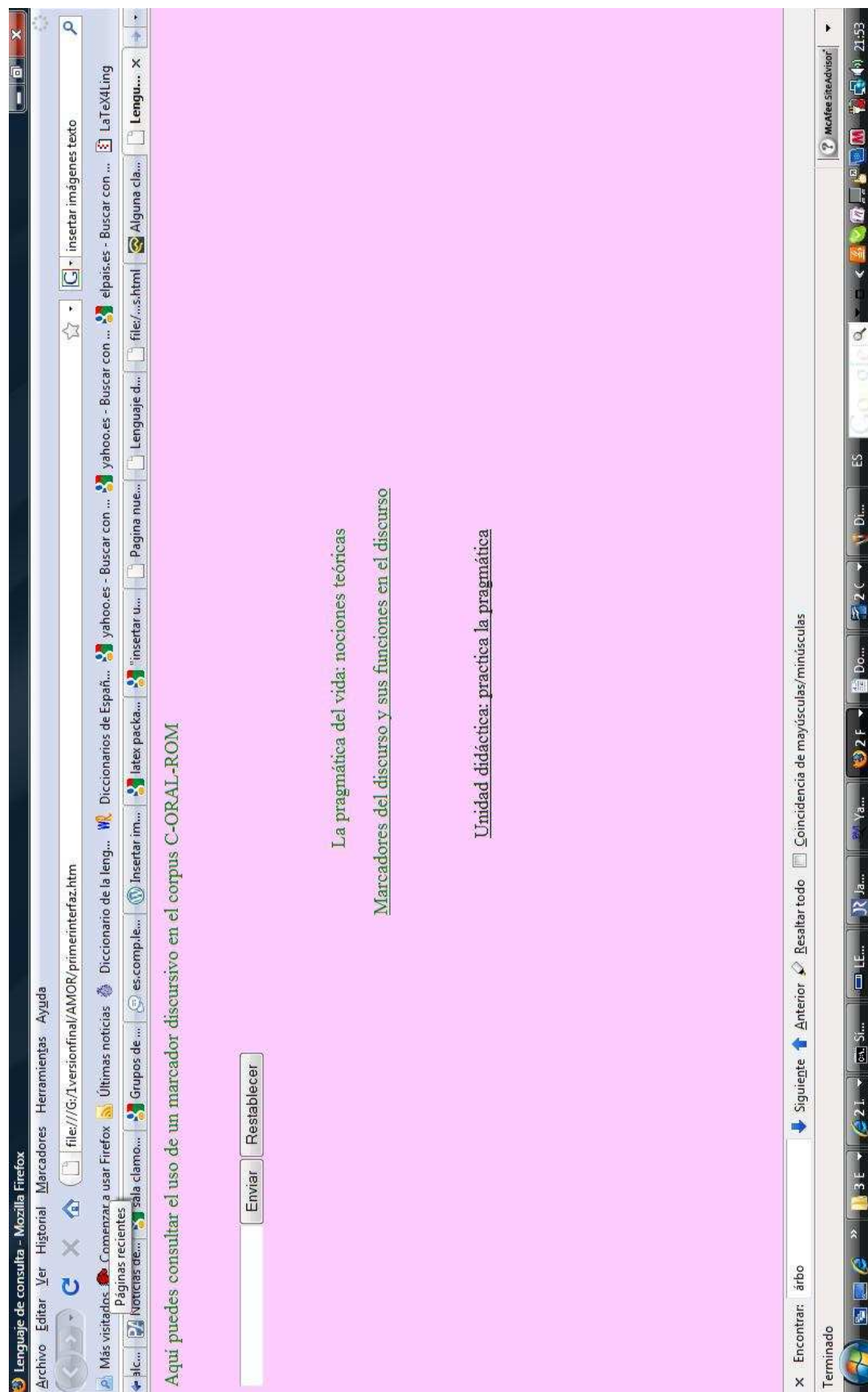


Figura 9.1: Primera página de Inter Alia

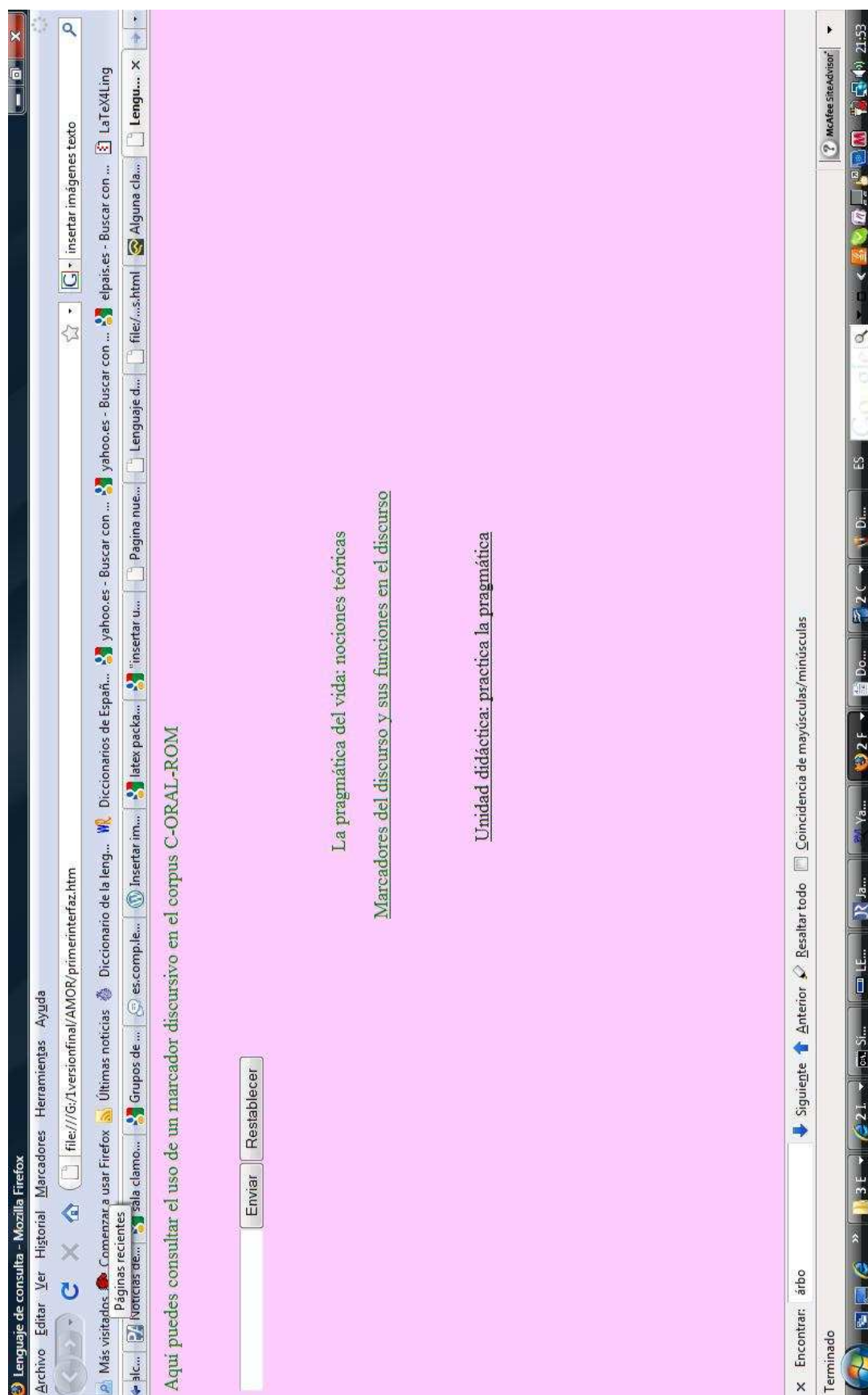


Figura 9.2: Segunda página de Inter Alia

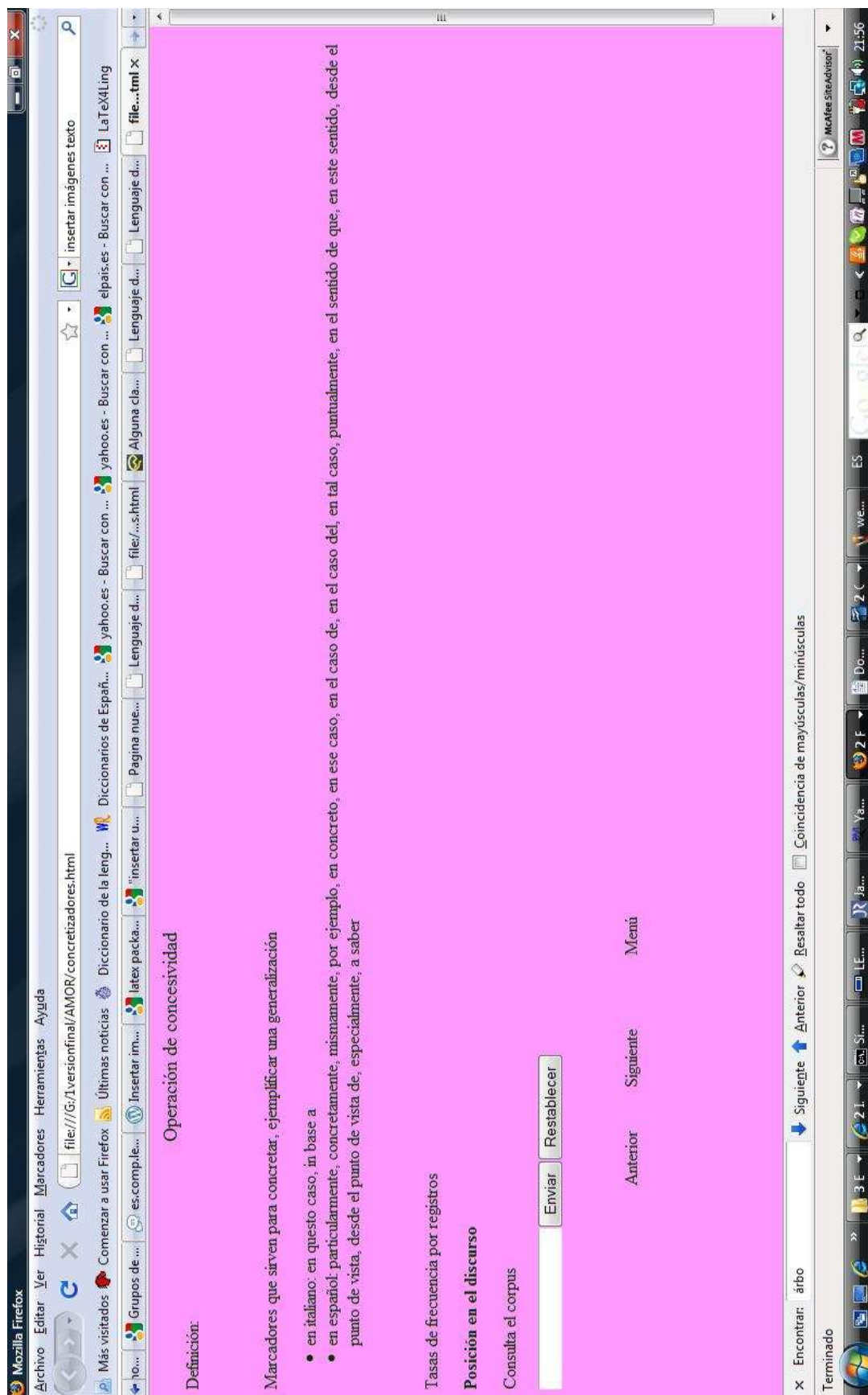


Figura 9.3: Tercera página de Inter Alia



Figura 9.4: Cuarta página de Inter Alia

9.4. Conclusiones

Llegamos al final de la segunda aplicación de un corpus anotado con información pragmática. En este capítulo hemos presentado las bases teóricas más importantes en la disciplina de la Enseñanza de español a alumnos extranjeros, y hemos diseñado una interfaz Web para introducir al alumno en el mundo de los marcadores del discurso y de los fenómenos lingüísticos que codifican, así como de la dimensión cognitiva, social y emocional que el uso de estas partículas conlleva en la interacción, atendiendo especialmente al cuidado de la imagen pública del alumno. Pero antes de pasar al último capítulo, me gustaría brevemente hacer algunas reflexiones críticas sobre el modelo funcionalista y la enseñanza de idiomas.

1. **Aprendizaje y funcionamiento de la mente:** La enseñanza de idiomas no debe desligarse del conocimiento sobre el modo en que funciona la mente. Después de treinta años de propuestas y modelos y teorías no sabemos casi nada de los procesos mentales cognitivos y afectivos que intervienen en el proceso de aprendizaje (BARALO, 1998). Tampoco encontramos reflexiones profundas y aplicaciones didácticas coherentes con dichas reflexiones en torno al tiempo, la resistencia, el inconsciente, las emociones, la actuación social o la autoestima. Y que aclararía cuestiones de enorme relevancia como la diferencia entre la adquisición y el aprendizaje.
2. **La naturaleza positiva del error:** Si bien es cierto que no debe cargarse al alumno emocionalmente con el error, una persona puede cometer un error en la tarea y el cerebro puede entender que ese error forma parte de la tarea, la plasticidad cerebral le permite juzgar ese error como correcto y el cerebro actuará de forma errónea se ha producido una plasticidad errónea (PUNSET, 2006). Con lo cual, el error no tiene una naturaleza positiva. Aunque parezca un tópico, debemos ser conscientes de los errores que cometemos; y el entorno debería indicarlo.
3. **El valor de la repetición en el aprendizaje:** Con respecto a cómo aprende el ser humano: la inteligencia humana es una interacción constante entre los datos mentales y los de la experiencia, cuanto más se repita una palabra más se refuerza en un lugar de la memoria de la mente. Con lo cual, es necesario convertir las reglas gramaticales y los contenidos en rutinas de pensamiento. Ya que hay una relación directamente proporcional entre velocidad de pensamiento y conocimiento interiorizado. Cuanto más

se repite algo, más se interioriza y pasa a ser parte del conocimiento subconsciente. Lo cual nos lleva a la siguiente crítica.

4. **Rekursividad frente a convencionalización:** El método comunicativo no favorece la recursividad. Al tiempo que no hace suficiente hincapié en la norma, ya que, cuando pensamos en otro idioma, puede haber una gran distancia entre el qué quiero decir y el qué se suele decir para expresar lo que yo quiere decir en esta situación comunicativa. Los estudiantes del árabe, que es una lengua semítica, reconocen la facilidad de la recursividad de las raíces para expresar campos semánticos. No debemos minusvalorar la gran utilidad de una regla para construir oraciones que expresan situaciones diferentes. Podríamos decir que son dos acercamientos al aprendizaje que reflejan dos componentes de la lengua necesarios y en constante tensión: recursividad y convencionalización. La recursividad ofrece posibilidades, pero la convencionalización escoge una, y esto es el famoso *se dice así*. Cuando se aprende una lengua se sobregenera, pero no se sabe si la posibilidad elegida es la que corresponde a la norma. Este es el juego de fuerzas entre la recursividad y la convencionalización. Así pues, un hablante no nativo puede generar *opinión del público* porque no sabe que para ello ya se ha fijado la colocación *opinión publica* para expresar lo que él quiere decir en realidad.

5. **Enseñanza de las palabras malsonantes.** Esta no es una crítica a los modelos funcionalistas sino a los profesores de lenguas extranjeras en general, basada en discusiones al respecto que he mantenido con diferentes profesores de lenguas a lo largo de mi vida. En dichas discusiones intento convencerles de que la mayor parte de nuestro sistema conceptual se estructura metafóricamente, conceptos tales como el amor, el tiempo o las emociones requieren una definición metafórica ya que dichos conceptos no están claramente estructurados. Un ejemplo: la palabra *cojones*, es un intensificador de la fuerza ilocutiva de los enunciados de uso cotidiano en el habla de Cádiz. Con muchísima frecuencia las afirmaciones acaban con esta palabra: *Que sí, cojones*, *No voy a tener razón, cojones*, *Que lo pongas ahí, cojones*, etc. Esto por no mencionar los famosos deícticos sociales *chocho* [shosho] y *picha* [pisha] de uso absolutamente generalizado en el gaditano coloquial, y que un profesor de español en esta zona debería estar obligado a explicar utilizando el marco de la metáfora.

Cerramos con esta última reflexión el capítulo nueve, y abrimos paso al último capítulo de la tesis, donde hablaremos de la relación entre Pragmática e Inteligencia Artificial.

Parte V

De la interpretación

Capítulo 10

Inteligencia artificial y Pragmática: perspectivas de investigación

Una tecnología suficientemente avanzada es indistinguible de lo mágico. Arthur C. Clarke.

Me gustaría comenzar este apartado apelando a la reflexión de la introducción en torno a la relación entre lenguaje natural y matemáticas. Creo que el debate fundamental en esta cuestión se centra en torno a los modelos matemáticos. A día de hoy, no estamos tan seguros de si las Matemáticas, como disciplina, vaya a aportar modelos matemáticos que sean aplicables al funcionamiento lenguaje natural durante la comunicación. La expresión de *cuello de botella* procedente del ámbito disciplinar de la Inteligencia Artificial se ha convertido ya, a fuerza de su uso, en un clásico dentro de los que están en contra de una Inteligencia Artificial fuerte, es decir, aquella que se plantea como reto científico la creación de una Inteligencia Artificial tomando como punto de referencia el funcionamiento de la mente. La implementación del lenguaje dentro de una estructura cognitiva no supone un problema baladí si escuchamos a los lingüistas, y dejamos de ver solo cadena de caracteres. Me complacería saber que el lector ha encontrado en este trabajo el significado real del gran reto que supone la representación del significado a partir solamente de la forma lingüística que la codifica.

Este capítulo junto con las conclusiones pertenece a la parte de la tesis denominada **De la Interpretación**. Lo que el lector encontrará aquí será un conjunto de reflexiones que ponen de manifiesto las carencias de la Lingüística Computacional y de la Inteligencia Artificial en sus modelos de lenguaje. Buena parte de estas carencias está causada por la ausencia de modelos de lenguaje que subyazcan a la implementación informática, lo cual nos remite a la posición que ocupa el lingüista, en particular, y la Lingüística, en general, tanto en las empresas como en la investigación pública dedicada a la Ingeniería Lingüística. Este tema es el punto de partida que arranca la argumentación presente en este capítulo,

donde se hablará de cómo, a la luz de los conocimientos que ha puesto de manifiesto la Pragmática, se deben reformular los planteamientos de tres parcelas típicas de esta parcela de investigación: la derivación automática de la estructura retórica de los textos, los sistemas de diálogo hombre-máquina y los sistemas de recuperación de información.

El 17 de Mayo del 2009, David Trueba hizo para la edición del domingo del periódico El País un reportaje sobre su amigo Pep Guardiola, a propósito de su triple éxito conseguido en la temporada de fútbol. En este documental, David Trueba, un intelectual al que en ningún caso se le puede atribuir una ideología de derechas, hizo la siguiente argumentación:

Viene de una familia muy humilde, pero son muy brillantes, muy honestos.

Ya hemos hablado en el capítulo tres y cuatro del significado del marcador discursivo *pero*. Dicho marcador contraargumenta, esto es, le confiere más peso argumentativo al enunciado que le sigue, volviendo falso el topos en el que se apoya el significado del enunciado anterior. En este caso en concreto, los topos son *los pobres son tontos y los pobres delinquen*. Esta asociación entre pobreza e inmoralidad y falta de inteligencia, como sabemos, es un lugar común creado bajo unos intereses ideológicos, ya que, desde un punto de vista científico, no se ha demostrado todavía la relación intrínseca y determinante que existiría entre ser pobre y delinquir y ser pobre y no ser inteligente. Del análisis de este enunciado se pueden sacar dos conclusiones. Una menos interesante de cara a nuestra tesis pero de obligada mención: el modo en que ciertas ideologías hacen de sus ideas lugares comunes que los periodistas de izquierdas asumen inconscientemente e inconscientemente se ven obligados a rebatir. La segunda conclusión que debemos mencionar aquí sí tiene trascendencia para los sistemas de generación y comprensión del lenguaje natural.

En primer lugar, es necesario señalar que lo que hace que unos enunciados aparezcan asociados a otros no son los marcadores del discurso sino las bases de conocimiento que están detrás. Los marcadores del discurso materializan esta relación. Hacen que dos enunciados entre los que en principio no tiene por qué haber una relación, como por ejemplo, entre ser pobre y ser brillante, se legitimen construyendo significado. En segundo lugar, hay que destacar que durante la sucesión de enunciados, lo que se entiende por discurso, ya sea escrito o hablado, forma parte integrante y determinante del conocimiento implícito en la que se apoyan las proposiciones que sí han sido verbalizadas. El lenguaje, como ya hemos señalado, está codificando solo parte de los procesos inferenciales que tienen lugar en la mente. Así pues, la transición del primer enunciado *vienen de una familia humilde* y el segundo *pero son muy brillantes, muy honestos* está justificada por un cálculo mental entre proposiciones explícitas y no explícitas.

Es por ello que la Lingüística Computacional y la IA deben convencerse definitivamente de la idea de que la forma lingüística no es suficiente, sino que hay que enriquecerla con fuentes de conocimiento. Ni las metodologías en Lingüística Computacional que van de la forma al significado ni las que van del significado a la forma funcionarán nunca, ya que no hay una relación biunívoca entre forma y significado, y lo que está determinando esta unión son factores de bases de datos organizadas en sistemas coherentes conformadas por topos cancelables.

Ya hemos comentado en el capítulo tres las iniciativas de D. Marcu para derivar automáticamente estructura retórica de los textos. Este investigador genera una representación en forma de árbol de la estructura de los textos en las que se explicita la relación semántica que guardan los enunciados allí presentes. Para ello se sirve de los marcadores del discurso, pero, en sus últimos estudios, ha conseguido explicitar relaciones semánticas no introducidas por estas partículas. Estas aproximaciones parten de la idea de que la forma lingüística en un texto es suficiente para adivinar la relación de significado entre sus partes; ya que se apoyan en la idea de que un texto es algo coherente. Sin embargo, la coherencia, como también se ha señalado es una cuestión de conocimientos y no una cuestión puramente formal. Para que una inteligencia artificial pudiera generar el enunciado con el que se ha comenzado este capítulo necesitaría algo más que unas reglas sintácticas, un lexicón y una instrucción sobre el significado de la conectiva *pero*. Necesitaría tener en su base de conocimientos dicho conocimiento: *Los pobres no son inteligentes*. Este conocimiento es falso, es una creencia, y por tanto es cancelable, tal y como indica Ducrot, los topos son cancelables, es decir, se pueden contraargumentar. Solo si la máquina tiene implementado que este conocimiento sirve, es útil, pero relativo, es cuando la máquina podrá generar contrarrestar el primer enunciado y cuestionar su valor de verdad con el segundo.

Las bases de conocimiento de un ser humano son subjetivas, limitadas pero guardan cierta coherencia; y están llenas de enunciados que representan conocimientos cancelables. El mayor exponente de dichos enunciados es el conocimiento codificado en los refranes, máximas, citas, etc. Estas bases de datos varían de un humano a otro incluso dentro de una misma cultura y están organizadas contextualmente. Así pues, pensemos en el conocimiento lingüístico de un carpintero sobre los tipos de madera, barniz, herramientas y muebles.

La IA se ha centrado mucho en la formalización del conocimiento contextual y en la creación de contextos para la realización de determinadas actividades. Sin embargo, y con esto volvemos al capítulo uno, el conocimiento que se debe implementar en el diseño de contextos debe incluir esta representación del mundo subjetiva que los hombres materializan en forma de normas. En lengua española, tenemos sistemas de diálogo que han incorporado conocimiento de la Pragmática introduciendo en su modelos de lenguaje algunos

marcadores del discurso, algunas formas de cortesía, el reconocimiento de formas típicas de la oralidad, y actos de habla (CALLE, 2005) (RODRIGNO y otros, 2002) (RODRÍGUEZ, 2004) (FERNÁNDEZ, 2000). No ponemos en duda el valor que representan estas iniciativas, sin embargo, es bastante obvio que de nuevo se aborda el problema en una dirección inadecuada. De la forma lingüística a la interpretación, y no de la interpretación a la forma lingüística. Dentro de los planteamientos que van de la forma lingüística a la interpretación, incluso aunque con un tope se pueden mejorar estos sistemas. La investigación que hemos hecho aquí puede enriquecer los futuros módulos de procesamiento discursivo que necesariamente tendrán que tener todos los sistemas de diálogo hombre-máquina. En el terreno del análisis de la señal, esta no solo debe segmentar los enunciados sino localizar todo tipo de partículas lingüísticas que digan algo sobre la posición del hablante con respecto a lo enunciado, por un lado, y sobre la relación semántica que guardan entre sí dichos enunciados. Con respecto a la primera de ellas, la localización de estas partículas, entre las que se encuentra, como categoría frecuente, la interjección, puede ayudar a revelar información sobre el estado emocional del hablante durante el desarrollo de la interacción comunicativa. Si al sistema se le enseña a reconocerlas como patrones ya conocidos (para lo cual habría que incluir un lexicón de marcadores) y después el analizador lingüístico les otorga un significado emocional (negativo o positivo) esta información evaluativa puede influir en los sistemas de decisión de acciones de la máquina. Así, si el hablante emite alguna expresión por ejemplo negativa, si la máquina la reconoce, puede plantearse la estrategia que tenía planeada, y saltar a otra subrutina. Lo mismo ocurre con la implementación del significado de ciertos marcadores discursivos. Por ejemplo, si en un sistema de diálogo de información sobre horarios de trenes, como INFOTREN, la máquina dice algo del tipo: Quiere usted salir el viernes a las diez y media, hable por favor. Y el hablante en lugar de decir sí, dice el marcador del discurso *digo*, como normalmente se responde a afirmaciones muy evidentes en el caso del habla del sur de España, la máquina debe interpretar que el valor de verdad del enunciado que ha emitido es verdadero, y debe pasar al siguiente paso del programa. Aún así tendremos que asistir a estudios de carácter crítico por parte de los pragmatistas que han evaluado sistemas de diálogo hombre-máquina y que han constatado cómo ninguno de ellos respetan las máximas de Grice ni el principio de cortesía (GONZÁLEZ-LEDESMA y MORENO, 2005) y (SAYGIN y CICEKLI, 2002). Los sistemas de diálogo que incluyen en su diseño conceptos procedentes de la Pragmática, como la estructura del discurso, la estructura de las conversaciones, el estudio de ciertas marcas de cortesía, o la inclusión de actos de habla no han explicitado hasta sus últimas consecuencias lo que Goffman llama el *frame*. Así pues, dentro de la formalización de los contextos, los sistemas de inteligencia artificial, tanto los de diálogo como los que se plantean retos mayores, como la integración

de información de otra naturaleza como por ejemplo la visual, deben primero modelar no tanto una idea de mundo, sino un escenario gobernado por reglas, donde los conocimientos, aunque sean pocos, estén conectados computacionalmente entre sí. Pongamos un ejemplo. Queremos diseñar un robot que atienda en un banco a clientes. Este es un robot muy sencillo, esto significa que sus fuentes de conocimiento solo están limitadas al contexto en el que va a actuar. El conocimiento que se le implemente tiene que estar conectado entre sí. Así por ejemplo, imaginemos que el módulo de reconocimiento de caras humanas está muy desarrollado, y es capaz de clasificar los sujetos en cuatro clases: femenino, masculino, joven y viejo. Esta información no es del todo relevante para el sistema de honoríficos del japonés. Con lo que si el robot hablara japonés utilizaría la forma neutra. Porque las categorías que son relevantes para un japonés a la hora de clasificar socialmente a una persona son demasiado finas para nuestra inteligencia artificial que, de momento, ve el mundo en cuatro categorías. Cuando una máquina enlace la imagen de un cliente y la designe con una forma de tratamiento: tú o usted, señor, señorita, caballero o señora, entonces estará creando significado. Al mismo tiempo, hay instrucciones que el robot debe conocer, como por ejemplo, el concepto de cola, esto es, no atender a un cliente que ha llegado después. Esta norma se puede implementar siempre que la percepción del robot esté hecha para localizar este tipo de situaciones. Así es como poco a poco se va creando una subjetividad que en este caso es muy limitada.

El diseño de un modelo de marco contextual, de las normas que lo rigen y de los actos de habla que van a predominar deben ser hechos por lingüistas, sociólogos y antropólogos que tengan la suficiente creatividad y capacidad de abstracción como para crear modelos de normas que guarden entre sí una relación coherente e implicativa, y que van a llenar los esquemas de razonamiento de la máquina.

Así pues, resumiendo lo dicho hasta aquí, debe haber una implementación del conocimiento, dicha implementación debe estar previamente modelada por un experto en la materia y debe ser contextual para que sea abarcable y limitada.

El último tema a tratar sobre los modelos de lenguaje y su relación con las bases de conocimiento está relacionado con el problema de las rutas, esto es, los caminos más rápidos para llegar a una solución.

Dentro del cerebro, el significado es el recorrido de una ruta neuronal. Cuantas más veces se haya hecho este recorrido más inmediata será la conexión, dichas rutas son contextuales, y la parte de la Lingüística Computacional que se dedica a organizar el conocimiento en ontologías está muy lejos de reproducir dichas rutas de información. Las ontologías no dan buen resultado, porque no está claro que esta estructuración del conocimiento lingüístico reproduzca las rutas por las que unas palabras se asocian a otras, por no hablar de las

rutas que hacen que unos enunciados se asocien a otros generando discurso, en el sentido más ideológico del término. La ruta que asocia *pobre* a *deshonestidad* no es una ontología sino un sistema de conocimientos basado en una percepción inter-subjetiva (esto es, compartida por los miembros de una comunidad) de la realidad.

Las teorías de la argumentación y del razonamiento de la IA se centran en la implementación de los patrones formales de estas operaciones mentales. Ya señalamos en el capítulo uno que la formalización del sentido común es un problema de contenido de los razonamientos y no de su forma lógica, asignatura que a día de hoy podemos considerar superada, como podemos considerar superado el problema de la generación de oraciones, gracias a la implementación del modelo de gramática procedente del generativismo. La forma lógica y programación de un razonamiento del tipo *Si bebes no conduzcas* ya no es un problema sin solución. Lo que es un problema es registrar todos los contenidos de los razonamientos que se toman en cuenta a la hora de planear una acción.

La Pragmática es un cálculo entre los diferentes factores contextuales antes de llevar a cabo un tipo de conducta lingüística o no lingüística. Este cálculo debe estar mediado por el coste y el beneficio social de dicho comportamiento. Así por ejemplo, cruzar la calle con el semáforo en verde es producto de un cálculo pragmático basado en reglas. Una conversación que comienza con un interés en lo personal y termina con una petición al interlocutor también está reflejando un cálculo social según el cual se intenta amortiguar el coste de la petición antecediendo actos de cortesía positiva. Si no se tiene en cuenta este conocimiento en los modelos de conducta de las máquinas seguiremos en el mismo punto en que nos encontramos ahora.

La Ciencia Cognitiva está muy interesada en el pensamiento meta, esto es, la reflexión sobre el propio pensamiento. Esta actividad puede ayudar a descubrir los principios formales que determinan el funcionamiento de la mente incluido el lenguaje. Sin embargo, con lo que tenemos que enfrentarnos cotidianamente es con datos e información estructurada, con las reglas ya materializadas, y sobre los cuales hay que hacer una labor de inducción para extraer modelos.

Aquí termina el capítulo diez. Podemos decir, resumiendo lo dicho hasta aquí, que si bien es difícil modelar la conducta general del ser humano, menos difícil es sin embargo modelar cada uno de los personajes que, siguiendo a Goffman, desempeña el ser humano en los diferentes contextos de su vida cotidiana según el rol social que desempeña. Así, si una máquina tuviera implementado el sistema de formas de tratamiento del japonés, sabría cómo tratarme solo después de que se diera un marco contextual y unas reglas que regulen el uso de una u otra forma en función del marco que tenga implementado la máquina. Pero solo un lingüista está formado académicamente para elaborar un modelo de cortesía para

un contexto determinado en función de los roles sociales, ya que es al lingüista el que le corresponde la reflexión entre forma lingüística y un significado contextual modelado previamente por un sociólogo o antropólogo.

Capítulo 11

Conclusiones finales y trabajo futuro

Hemos llegado al final de la investigación. Vamos a estructurar la última sección de este trabajo en tres partes. En la primera de ellas haremos un resumen de los contenidos de la tesis y destacaremos el impacto o novedad de los conocimientos aquí presentados. En la segunda parte, expondremos lo que a mi juicio son los puntos débiles o críticos de esta investigación, y que deben tener una continuación en el futuro. En tercer y último lugar, cerraremos este trabajo con una reflexión general en torno a la relación mente, lenguaje y realidad, incidiendo en la importancia de la ficción y retomando algunas ideas de la teoría de la polifonía.

11.1. Resumen y novedad del trabajo

Comencemos con un breve resumen de la investigación que se ha llevado a cabo en esta tesis. En primer lugar, nos gustaría destacar que este trabajo se ha ocupado fundamentalmente del problema de cómo gestionar los marcadores discursivos en un corpus concreto, C-Oral-Rom, en todas y cada una de sus etapas, desde la más puramente teórica hasta su aplicación a otras parcelas de conocimiento, pasando por el imprescindible trabajo computacional que es necesario hacer cuando se debe gestionar una cantidad importante de datos. Para ello, hemos atendido a los marcos teóricos más sobresalientes de esta disciplina, hemos visto qué resultados han dado la aplicación de esos enfoques al estudio del significado de los marcadores del discurso, y con esta información hemos diseñado un esquema de anotación basados en rasgos pragmáticos. Este modelo de anotación está destinado al etiquetado de partículas que hacen referencia a procesos de la enunciación del discurso, a los valores de verdad de los enunciados y al modo en que los enunciados empiezan a influirse unos a otros en términos de valores de verdad. Como se puede observar, este tipo de información, como por ejemplo, palabras que codifiquen atenuación, no tienen nada que ver

con la información que la Lingüística de Corpus nos ha ofrecido hasta el momento de los textos, como la información que nos habla de las categorías gramaticales de las palabras, las que diferencian eventos y sus argumentos, o las que explicitan funciones sintácticas. Estamos dando cuenta aquí de un nivel de la lengua, propio de la enunciación, que si bien no soluciona el problema del significado de los enunciados y la relación entre ellos, sí se considera un paso necesario a la hora de plantearse la resolución computacional de este problema. Más concretamente, Pragmatext, nuestro modelo de anotación, trata de la presencia de los siguientes fenómenos lingüísticos: emociones, deixis, certeza, evidencialidad, metáfora, convencionalización y operaciones del discurso. Cada fenómeno ocupa un rasgo y aporta un valor sobre el significado final del enunciado. Además, el hecho de colocarlos como valores separados da la posibilidad de reflejar la influencia que tienen unos rasgos sobre otros de una manera discreta. Una vez que hemos diseñado el esquema de anotación, hemos desarrollado estrategias para introducir dicha información en los textos de la manera más automática posible. Para lo cual hemos diseñado el etiquetador de marcadores discursivos, DiMaSTagger. Una vez que todos los marcadores del discurso tienen en los textos su etiqueta correspondiente hemos presentado dos aplicaciones en las que este nivel de anotación puede ser útil a la hora de recuperar información lingüística en los textos. La primera de ellas ha sido en el terreno de la Lingüística Cuantitativa, donde hemos analizado la frecuencia de los diferentes marcadores en relación a los géneros discursivos. En la segunda, hemos realizado una interfaz Web para que el estudiante ELE pueda educar, por medio de los contenidos presentados, su sensibilidad hacia la dimensión pragmática de la lengua a través del estudio de los marcadores del discurso del español oral. Finalmente, en la última parte, denominada **De la interpretación** hemos reflexionado sobre la relación entre Pragmática, Lingüística Computacional e Inteligencia Artificial, haciendo hincapié en los desafíos que representa para estas disciplinas la implementación de este conocimiento.

Así pues, para resumir las novedades de la tesis, hemos intentado contribuir a diferentes parcelas de conocimiento intentando salvar las carencias que dichas parcelas presentan en algunos aspectos. A nivel teórico-descriptivo, dentro de la Pragmática Lingüística, esperamos haber contribuido con este estudio a la recolección y sistematización de esta clase de palabras. En lo que respecta a la Lingüística de Corpus, hemos diseñado un modelo de anotación pragmática que intenta recoger con amplitud los fenómenos de los cuales esta parcela se ha ocupado e intentado salvar la carencia de modelos de anotación en este nivel que encontramos en este terreno de investigación. En lo que respecta a la parte computacional, no encontramos etiquetadores de marcadores que reconozcan y clasifiquen estas partículas en lengua española. En penúltimo lugar, dentro de las aplicaciones, esperamos haber abierto una veta en la Lingüística Cuantitativa de los estudios en lengua española

de carácter empírico, en lo que se refiere a la aplicación de las técnicas de la Lingüística Computacional al estudio de la frecuencia de los marcadores del discurso en función del género discursivo. Finalmente, en la parte dedicada a la generación automática de lenguaje natural, disciplina que entraría dentro de la Inteligencia Artificial, esperamos haber contribuido a nivel teórico con nuestras reflexiones sobre el lugar que la Pragmática debería ocupar dentro de este campo.

Establecidos pues los puntos novedosos de la tesis, pasemos a explicar aquellos puntos susceptibles de crítica y de mejora en investigaciones futuras.

11.2. Puntos críticos y trabajo futuro

Forma parte del trabajo futuro, y por ende, de las limitaciones de la investigación realizada aquí, la integración del sistema de anotación pragmática dentro de una investigación mayor. En concreto, el Laboratorio de Lingüística Informática deberá en el futuro unificar todos sus esquemas de anotación, y todos sus etiquetadores (el categorial y el pragmático). Además, para completar el etiquetado del C-Oral-Rom falta el nivel sintáctico. La interfaz sintaxis pragmática no es una cuestión superficial. Pensemos en las posibles realizaciones discursivas del sintagma *qué fuerte* en estos contextos.

1. Qué fuerte no me lo puedo creer.
2. Qué fuerte ha sonado.

Aquí nos encontramos con dos análisis sintácticos de la oración. Si el análisis computacional se hace de manera secuencial ¿qué módulo pasaríamos antes? ¿el que gestiona la información sintáctica o el que se ocupa del nivel pragmático¹? ¿Cómo podemos solucionar el problema de diferenciar entre las estructuras o secuencias de palabras que en un contexto tienen un significado composicional (esto es, que el significado sea la suma de sus constituyentes) y las que poseen un significado convencional, fraseológico y en algunas ocasiones, figurativo? Amplíese este problema a los casos de todas las unidades fraseológicas incluidas las colocaciones, y no solo al de los marcadores del discurso. ¿Cómo distinguir a priori una colocación de sintagma? Las reglas sintácticas no solo combinan unas palabras con otras sino construcciones de diferente tamaño y naturaleza. La Lingüística Computacional no puede hacer oídos sordos del hecho de que lo que detiene la recursividad es la convención. Dejemos esta cuestión apuntada aquí para ser planteada en el futuro

¹Reflexión fruto de la conversación con el lingüista computacional Fernando Sánchez León, en la RAE.

cuando un texto tenga que pasar por todos los niveles de etiquetado descritos, cada uno de ellos aportando una información sobre las construcciones que allí se encuentran.

Respecto del modelo de anotación, debemos mejorar la tipología de actos de habla y la de la evidencialidad. Ya que ha sido elaborada *ad hoc* a medida que íbamos etiquetando las palabras. Además, es necesario contrastar la clasificación semántica de algunas unidades discursivas, que pueden presentar otros significados en otros corpus. Por último, la crítica más importante del modelo de anotación es que no se ha reflejado la escalaridad o gradualidad de estos fenómenos. Se ha hecho una conceptualización de la expresión de la certeza discreta, esto es, hemos indicado si hay atenuación o intensificación de la certeza, pero como sabemos la certeza se estructura en grados, grados de los que el modelo de anotación no da cuenta.

El tercer punto débil se encuentra en el etiquetador. Ya que la identificación de marcadores del discurso en el corpus también se ha hecho *ad hoc*. Esto dificultaría las labores automáticas cuando cambiáramos de corpus. Además, las reglas contextuales diseñadas solo funcionan para corpus orales, y aunque no es difícil adaptarlas a los corpus escritos, solo hay que sustituir las marcas prosódicas por los signos de puntuación, ignoramos su comportamiento en términos de ambigüedad, aunque también es cierto que en la parte formal del corpus existen muchas transcripciones que se apoyan en texto escrito como, por ejemplo, las conferencias y los sermones. Por último, hay que perfeccionar las reglas de desambiguación oral. Por ejemplo, la regla 3.4, constituida por Foco+Transcription+Mark MD coge contextos del tipo:

*ANA: te molesta el aparato joder // o sea ...

donde el marcador *o sea* ya pertenece al otro enunciado y no tiene influencia sobre la desambiguación de la partícula anterior. Se puede ser más fino, y hacer diferencias entre una barra y dos barras en el diseño de las reglas. Por otra parte, y ya para terminar, en el caso de haber tenido todos los niveles de etiquetados concentrados en un mismo texto se podría haber conjugado la información categorial con la prosódica en la elaboración de las reglas. De esta manera, el enunciado

*ANA: si no pasa nada

hubiera sido mucho más fácil de desambiguar. En el futuro, el Laboratorio de Lingüística Informática deberá ocuparse de integrar toda la información a la hora de hacer las reglas de desambiguación.

Forma parte del trabajo futuro, integrar el conocimiento aquí generado dentro de lo que

sería el módulo dedicado al procesamiento pragmático de un sistema de diálogo hombre-máquina.

Una vez terminados de enumerar algunos de los puntos débiles de esta investigación, pasemos a la reflexión final de este trabajo.

11.3. Las voces de la razón: retos para la relación entre mente y lenguaje.

En primer lugar, me gustaría destacar que la dicotomía entre significado cognitivo y significado socio-cultural es tan útil metodológicamente como falsa en la realidad, ya que en el fondo la cultura forma parte de la cognición. Comenzando por esta última, la Teoría de la Relevancia, nos habla de qué ocurre con el significado y la forma desde una perspectiva si se quiere más psicológica. En este sentido, Sperber y Wilson, pero también algunos pragmatistas como Grice, inciden en que el significado es el resultado de una inferencia. Lo que me interesa señalar aquí es que el significado de una forma lingüística aflora en el cerebro siempre dentro de un contexto que será el que marcará, evocará su significado más relevante. La forma lingüística, ya lo dice tanto la teoría de la argumentación como la de la relevancia, convoca el significado. En el capítulo cinco ya hemos explicado cómo diferentes valores pueden convivir en un mismo marcador del discurso, y cómo algunos de ellos conviven simultáneamente y cobran diferente grado de relevancia según el contexto de uso. La relevancia es un principio que funciona también a nivel de significado. Este proceso se puede observar claramente en la creación de términos nuevos, como por ejemplo, el término *nativos digitales*, el cual es el resultado de una operación de síntesis basado en los rasgos más relevantes de la realidad que se ha querido designar. Además cada vez que las formas lingüísticas son usadas en el tiempo y en el espacio se producen procesos de atribución de significado, y de construcción de otros nuevos. Este hecho se puede ver en los procesos de resignificación de los símbolos religiosos a lo largo de la evolución de las religiones. Un símbolo a lo largo de su historia se ve sometido un proceso de atribuciones de significado que cambian según los contextos. Entraríamos pues en la dimensión social de los significados.

Hasta ahora, hemos puesto los ejemplos de atribución de significado a nivel de palabra. Pero es apropiado recordar aquí que los procesos de atribución de significado se producen no solo en este nivel sino también en el de las unidades fraseológicas y en los del enunciado. No deja de ser relativamente curioso que la recursividad haya sido enunciada como una propiedad de la sintaxis cuando en realidad es una propiedad innata al lenguaje como lo

es, en último término, a la naturaleza del universo. La mente es recursiva, como cualquier sistema biológico, y en ella la información se recrea constantemente.

La gran aportación de la Pragmática al mundo del conocimiento es el énfasis que ha puesto en señalar que los enunciados están ligados unos a otros, que sus significados se enlazan con el lazo de la inferencia mental y que los marcadores del discursos nos ayudan a hacer restricciones sobre la interpretación de los enunciados anteriores. Así, en *Si bebes, no conduzcas*, nos encontramos con dos enunciados cuya relación no puede definirse en principio como lógica, puesto que no hay ninguna relación a priori entre conducir y beber; pero a través de la partícula *si* dicha relación se vuelve lógica. Es por ello que los marcadores del discurso deben estar representados formalmente, porque detrás de esta relación semántica hay una concepción del mundo que se ha denominado sentido común. Los relevantistas han dicho que los marcadores son piezas con un significado procedimental no representacional. En este trabajo hemos podido observar, a través del estudio de los significados de los marcadores de C-Oral-Rom, que los marcadores discursivos codifican información de muchas clases, procedentes de fenómenos lingüísticos de diversa naturaleza. Pero, sin embargo, ninguno de estos se considera en la representación lógica de enunciados, ni en la semántica proposicional, porque esta disciplina se ha centrado fundamentalmente en la representación del significado oracional, esto es, el significado como resultado de los elementos que componen la oración. Es necesario subir un nivel más y empezar a crear sistemas formales que reflejen no solo que los enunciados luego tienen un significado total, sino que además este significado está constreñido por la relación semántica que comparta con los enunciados que le rodean y con las proposiciones implícitas. Así es como se expresa el pensamiento consciente, en su fluir cotidiano.

Actualmente nadie duda de que el conocimiento del cerebro y su funcionamiento es con creces superior al de lo que se sabía de él en siglos anteriores. El programa marcado dentro de las Ciencias Cognitivas siempre se ha caracterizado por este afán de integrar las investigaciones procedentes de diferentes disciplinas (Lingüística, Neurología, Filosofía, IA, etc.). No obstante, muchos de estos marcos teóricos están muy lejos de ser compatibles entre sí. No solo porque utilicen terminología diferente, sino también porque los acercamientos teóricos, los modelos, conceptos y herramientas aluden a sistemas de clasificación diferente de un mismo objeto. Además, existe una brecha bastante grande por el momento entre aquellos cognitivistas que creen en un modelo computacional de la mente, y aquellos que abogan por la absoluta superación de esta metáfora, que consideran errada, ya que los últimos descubrimientos sobre el cerebro y su funcionamiento confirman que la mente es analógica, metafórica desde su concepción, y que opera mapeando unos conocimientos

con otros. El origen metafórico del lenguaje avalaría este funcionamiento básico (RATEY, 2002).

Los marcadores del discurso en este sentido nos hablarían de la sintaxis de los enunciados. Si bien no estoy insinuando que se puedan definir reglas para la construcción de textos, al igual que se han descrito reglas para construir oraciones, es justo señalar que hay conjuntos de textos que comparten rasgos estructurales, patrones. Las entrevistas, por ejemplo, comparten un patrón formal, las conversaciones telefónicas también, al igual que las cuatro mil novelas de Corín Tellado, o las partituras de Bach, tal y como el hijo del premio Nobel de física D. Hofstadter nos lo hace entender en su obra *Godel, Escher, Bach, un eterno grácil bucle*. Es en este sentido cuando definiendo la recursividad no del lenguaje sino también del pensamiento.

En la Expo de Zaragoza del 2008 había un experimento en el que las moléculas del agua cambiaban su forma en función de las palabras que proferían visitantes del pabellón. El lenguaje es un producto de la naturaleza y debe tener una explicación en todos los niveles de descripción incluido aquel que lo considera una energía que participa de la interacción con el universo. Citemos las palabras de Penrose en su obra *La mente del emperador* dentro del capítulo dedicado a explicar la relación entre consciencia y mente:

Sabemos que en el nivel submicroscópico de las cosas las leyes cuánticas son válidas; pero en el nivel de las bolas de críquet es la física clásica la que vale. Sostendré que en algún lugar intermedio necesitamos comprender la nueva ley para ver cómo el mundo cuántico enlaza con el clásico. ¡Creo también que necesitamos esta nueva ley si queremos conocer alguna vez las mentes! Por todo esto pienso que debemos buscar nuevas claves.

El lenguaje es la parte más accesible de la mente humana; todos queremos saber cosas sobre el lenguaje porque esperamos que todos estos conocimientos nos lleven entender mejor la mente. Dentro de la Lingüística parece que se han levantado dos bandos de reconciliación imposible. Podríamos hablar, por un lado, de aquellos que se preocupan por lo interno, dentro de los cuales podríamos denominar a los investigadores que realizan su labor dentro del generativismo y la pragmática cognitiva; y, por otro, de aquellos más preocupados por lo que de una manera burda podríamos denominar la actuación, y que a fuerza de estudiar las producciones lingüísticas en un medio real han llegado a ciertas dimensiones del lenguaje en las que el generativismo o el estructuralismo no se habían interesado en su justa medida. Uno de los fenómenos que han puesto de relevancia los estudiosos de los textos, tanto Ducrot con su teoría de la Polifonía como Bajtin en el terreno de la crítica literaria

rusa, es que las producciones lingüísticas son juegos de voces que representan diferentes posturas argumentativas (REYES, 1990). Pero es de justicia señalar aquí que es la mente la que es polifónica y no el lenguaje. En el capítulo 4 definimos la argumentación con la frase *dale a A una razón para pensar C*. Este A siempre es una representación mental. El pensamiento que se lleva a cabo a través del lenguaje interior es un acto de argumentación para llegar a la verdad y generar más conocimiento dentro de la mente. Esto no podría darse si constantemente no estuviéramos dudando sobre determinadas verdades y encontrando argumentos para apoyar unas y desbancar otras. Este proceso de búsqueda de la verdad se realiza de manera dialógica, donde cada voz representa una postura discursiva. La descripción de los esquemas de razonamiento en interacción, donde cada razonamiento remite a presupuestos diferentes, es solo una parte de lo que la IA tiene que empezar a considerar si quiere que sus redes neuronales artificiales dialoguen consigo mismas.

Se suele decir que la investigación en Inteligencia Artificial debe estar contextualizada a la realización de tareas que impidan al ser humano pensar de una forma creativa. Es muy interesante observar el modo en que se integran dichas afirmaciones en sistemas económicos que impiden al hombre precisamente pensar de manera creativa, y que lo convierten en mano de obra muy fácil de sustituir por una máquina a poco que implementemos en ella ciertos marcos contextuales de conocimiento, sobre todo en lo que a los puestos de trabajo más sencillos o rutinarios se refiere. Si el investigador en IA tiene una ética no neoliberal dejará sencillamente de investigar en este campo, como lo hizo el autor del programa Eliza, al ver cómo el resultado de sus investigaciones estaba destinado a inventar un robot soldado que continuara con la carrera colonialista de los EEUU en el mundo, antes de la llegada del presidente actual. Lo cierto es que siempre me he planteado qué clase de investigador sería el que inventó la máquina que obliga al personaje de Charlie Chaplin a comer en la película *Tiempos Modernos*. La llegada de robots que sirvan para ocuparse de cosas muy concretas será visto como un cambio positivo cuando el conjunto de la sociedad se sienta beneficiado de ella.

Parte VI

De las fuentes de conocimiento

Bibliografía

- ACÍN, E.: «Por cierto, a propósito y otros digresivos». En: P. GÓMEZ; P. CARBONERO y M. CASADO (Eds.), *Lengua y discurso : estudios dedicados al profesor Vidal Lamíquiz*, pp. 59–72. Arco/Libros, Madrid, 1999.
- ALCÁNTARA, M.: *Introducción al análisis de estructuras lingüísticas en corpus. Aproximación semántica*. UAM, Madrid, 2007.
- ALCÁNTARA, M. y BERTOMEU, N.: «Ellipsis in Spontaneous Spoken Language». *Workshop Cross-modular Approaches to Ellipsis, ESSLLI, Edimburgo (Escocia)*, 2005.
- ALLEN, J. y CORE, M.: «Draft of DAMSL: Dialog Act Markup in Several Layers». <http://www.cs.rochester.edu/research/cisd/resources/damsl/>, 1997.
- ALONSO, A.: «Las construcciones exclamativas. La interjección y las expresiones vocativas». En: I. BOSQUE y V. DEMONTE (Eds.), *Gramática descriptiva de la lengua española*, volumen 3, pp. 3993–4050. Espasa-Calpe, Madrid, 1999.
- ALONSO, L.: *Representing discourse for automatic text summarization via shallow NLP techniques*. Tesis doctoral, Universidad de Barcelona, 2005.
- ALONSO, L.; CASTELLÓN, I.; GIBERT, K. y PADRÓ, L.: «Discourse marker characterisation via clustering : extrapolation from supervised to unsupervised corpora». *Procesamiento del lenguaje natural*, 2002a, **29**, pp. 223–230.
- ALONSO, L.; CASTELLÓN, I. y PADRÓ, L.: «Lexicón Computacional de Marcadores del Discurso». *SEPLN, XVIII Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural*, 2002b, **29**, pp. 239–246.
- ALONSO, L.; SHIH, J.; CASTELLÓ, I. y PADRÓ, L.: «An analytic account of discourse markers for shallow NLP». *The Meaning and Implementation of Discourse Particles , whorkshop held as part of the Fifteenth European Summer School in Logic, Language and Information, ESSLLI'03, 18-19 August, Vienna, Austria*, 2003.

- ALONSO, M. C.: «La motivación afectiva y social como recurso para la expresión oral en el aula L2». En: A. MARTÍNEZ; J. A DE MOLINA y P. BARROS (Eds.), *II Jornadas sobre aspectos enseñanza del español como lengua extranjera*, pp. 17–22, 1993.
- ALVAREZ, M. y GINORIA, M.: «Aproximación al concepto de presuposición». *Anuario de Estudios Filológicos*, 1996, **19**, pp. 27–37.
- AMEKA, F.: «The meaning of phatic and conative interjections». *Journal of Pragmatics*, 1992, **18**, pp. 245–271.
- ANDERSON, L. M.; FISTER, A.; LEE, B.; TARDIA, L. y WANG, D.: «On the types and frequency of meta-language in conversation: A preliminary report». *14th Annual Meeting of the Society for Text and Discourse, Chicago*, 2004.
- ANSCOMBRE, J. C. y DUCROT, O.: *La argumentación en la lengua*. Gredos, Madrid, 1994ª edición, 1983.
- ARTIGAS, M.: *Filosofía de la ciencia*. Universidad de Navarra, Navarra, 1999.
- BACH, C.: «Los marcadores de reformulación como localizadores de zonas discursivas relevantes en el discurso especializado». *Debate terminológico, revista electrónica.*, 2005, **1**.
- BARALO, M.: *Teorías de la adquisición de lenguas extranjeras y su aplicación a la enseñanza del español*. Colección Experto en Enseñanza del Español como Lengua Extranjera, Fundación Antonio de Nebrija, Madrid, 1998.
- BEAU-FLY, J. (Ed.): *Estrategias para enseñar a aprender : un enfoque cognitivo para todas las areas y niveles*. Aique, Buenos Aires, 1995.
- BERBER, T.: «A Program for Identifying Metaphor Candidates in Corpora.» *Corpus Linguistics Conference Proceedings. Birmingham*, 2007.
- BERGER, D.; REITTER, D. y STEDE, M.: «XML/XSL in the Dictionary: The Case of Discourse Markers». *Proceedings of the 2nd Workshop on NLP and XML (NLPXML)*, 2002.
- BERMÚDEZ, F.M.: *Evidencialidad. La codificación lingüística del punto de vista*. Tesis doctoral, Universidad de Estocolmo, 2005.

- BERMEJO, M. L. y CASADO, J.: «Consideraciones didácticas sobre la enseñanza de lenguas extranjeras asistida por ordenador». *Didáctica (Lengua y literatura)*, 2000, **12**, pp. 67–89.
- BLAKEMORE, D.: *Semantic Constraints on Relevance*. Basil Blackwell, Oxford, 1987.
- : «¿Are apposition markers discourse markers?» *Journal of Linguistics*, 1996, **32**, pp. 325–347.
- BONILLA, S.: «Información y relevancia: una hipótesis acerca de cómo los seres humanos procesamos la información lingüística». *Revista española de documentación científica. CSIC*, 1996, **14(9)**, pp. 392–410.
- BOU-FRANCH, P. y GARCÉS, P.: «La presentación de la imagen en conversación entre hablantes nativos y no nativos del inglés». *Pragmalingüística*, 1994, **2**, pp. 37–61.
- BRIZ, A.: «Diccionario de partículas discursivas del español», 2008.
<http://textodigital.com/PDDPD>
- BRIZ, A. y HIDALGO, A.: «Conectores pragmáticos y estructura de la conversación». En: *Los marcadores del discurso. Teoría y análisis*, pp. 121–143. Arco/Libros, Madrid, 1998.
- BUNT, H. y BLACK, W.: «The ABC of computational pragmatics». En: H. BUNT y W. BLACK (Eds.), *Computational pragmatics: Abduction, belief and context*, John Benjamins, Amsterdam, 2000.
- BURGOS, M.: «Formalismo y funcionalismo: Concepciones teóricas y metodológicas divergentes dentro de la lingüística», 2008.
www.unm.edu/~fmburgos/manuelwepage/FormalismoFuncionalismo.pdf
- CACCIARI, C. (Ed.): *Similarity in Language, Thought and Perception*. Brepols Publishers, Belgium, 1995. Book reviews..
- CAFFI, C. y JANNEY, R. W.: «Towards a pragmatics of emotive communication». *Journal of Pragmatics*, 1994, **22(3)**, p. 325.
- CALLE, J.: *Interacción natural mediante procesamiento intencional: modelo de hilos de diálogo*. Tesis doctoral, UPM, Madrid, 2005.

- CAMPILLOS, L.; GOZALO, P.; MORENO, A. y GUIRAO, J.M.: «Exploiting a spoken corpus in language teaching/learning: an advance web/based tool». *Proceedings of 4th Corpus Linguistics Conference, University of Birmingham, 27-30 July, 2007*.
- CARLSON, L.; MARCU, D. y OKUROWSKI, M. E.: «Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory». En: *CURRENT DIRECTIONS IN DISCOURSE AND DIALOGUE*, pp. 85–112. Kluwer Academic Publishers, Amsterdam, 2001.
- CARRASCO, A.: «Revisión y evaluación del modelo de cortesía de Brown y Levinson». *Pragmalingüística*, 1999, **7**, pp. 1–44.
- CASADO, M.: *Introducción a la gramática del texto del español*. Arco/Libros, Madrid, 1993.
- : «Lingüística del texto y marcadores del discurso». En: M^a A. MARTÍN y E. MONTOLÍO (Eds.), *Los marcadores del discurso. Teoría y análisis*, pp. 55–71. Arco/Libros, Madrid, 1998.
- CASTILLO, M^a. A.: «Conocimiento Cultural en la Adquisición de la L2: la Fraseología». En: M. PÉREZ y J. COLOMA (Eds.), *El Español, Lengua del Mestizaje y la Interculturalidad: Actas del XIII Congreso Internacional de la Asociación para la Enseñanza del Español Como Lengua Extranjera (Asele). Asociación para la Enseñanza del Español Como Lengua Extranjera. Murcia*, volumen 13, pp. 209–216, 2003.
- CERDÁ, R.: «Dispersión y solapamiento en pragmática». *Revista española de lingüística*, 1995, **25**(2), pp. 271–296.
- CHOWN, M.: *Quantum Theory cannot hurt you. A guide to the universe*. Joseph Henry Press, Washington, 2006.
- CORPAS, G.: *Manual de fraseología española*. Gredos, Madrid, 1996.
- CORTÉS, L.: «Marcadores del discurso y análisis cuantitativo». En: M. MARTÍN y E. MONTOLÍO (Eds.), *Los marcadores del discurso. Teoría y análisis*, Arco/Libros, Madrid, 1988.
- : «Bibliografía: Marcadores del discurso I». *Español Actual: Revista de español vivo*, 1995a, **63**, pp. 63–82.
- : «Bibliografía: marcadores del discurso II». *Español actual. Revista del español vivo*, 1995b, **64**, pp. 75–94.

- : «1990-1999. Estudio de los distintos tipos de discurso oral». En: *Los estudios del español hablado entre 1950 y 1999*, Arco/Libros, Madrid, 2002.
- CORTÉS, L. y CAMACHO, M.: *Unidades de segmentación y marcadores del discurso*. Arco/Libros, Madrid, 2005.
- COTS I CAMINOS, J. M.; COTS, J.M.; ARMENGOL, L.; ARNÓ, E. y LLURDA, E.: *La conciencia lingüística en la enseñanza de lenguas*. Cercle de Lingüística Aplicada, Lleida, 2007.
- CRESTI, E. y MONEGLIA, M. (Eds.): *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. John Benjamins, Amsterdam, 2005.
- CUENCA, J. y HILFERTI, J.: *Introducción a la Lingüística Cognitiva*. Ariel, Barcelona, 1999.
- CUETO, N. y LÓPEZ, M. J.: *La interjección. Semántica y Pragmática*. Arco/Libros, Madrid, 2003.
- CURTO, N.; GONZÁLEZ-LEDESMA, A.; NICOLÁS, C. y VALVERDE, A.: «Diseño de una unidad didáctica con corpus del español oral ¿y si tú fueras el profesor?» En: C. NICOLÁS (Ed.), *Ricerche sul Corpus del parlato romanzo C-ORAL-ROM. Studi linguistici e applicazioni didattiche per linsegnamento di L2*, Firenze University Press, Florencia, 2007.
- DALGLEISH, T y POWER, M. (Eds.): *Handbook of Cognition and Emotion*. Wiley, New York, 1999.
- DEVILLER, L.; MARTIN, J. C. y ABRILIAN, S.: «Annotation of Emotions in Real-Life Video Interviews: variability between coders». *5th Int. Conf. Language Resources and Evaluation*, 2006.
- DIJK, T.A.: «Pragmatic connectives». *Journal of Pragmatics*, 2000, **32(6)**, pp. 687–707.
- DURANTI, A.: *Antropología lingüística*. Cambridge University Press, Cambridge, 2000ª edición, 1997.
- EDITORIAL: «Focus on issue formal and philosophical aspects of pragmatics». *Journal of Pragmatics*, 2007, **39**, pp. 787–791.
- ESCANDELL, V.: *Introducción a la Pragmática*. Ariel, Barcelona, 1996.

- : «Norms and Principles: putting Social and Cognitive Pragmatics together». En: R. MÁRQUEZ-REITER y M. E. PLACENCIA (Eds.), *Current Trends en the Pragmatics of Spanish*, John Benjamins, Amsterdam, 2004.
- ESULI, A. ; SEBASTIANI F. y URCIUOLI, I.: «Annotating Expressions of Opinion and Emotion in the Italian Content Annotation Bank». En: *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08). European Language Resources Association (ELRA)*, Marrakech, Marruecos, 2008.
- FAJARDO, L. A.: «La metáfora como proceso cognitivo». *Forma y Función*, 2006, **19**, pp. 47–55.
- FERNÁNDEZ, G.: *Un modelo para la especificación lingüística y la gestión computacional de diálogos hombre-máquina mediante instrucciones expresadas en lenguaje natural*. Tesis doctoral, Universidad de Sevilla, 2000.
- FICHER, K. y BRANDT-POOK, H.: «Automatic Disambiguation of Discourse Particles», 1998, pp. 107–113.
- FISCHER, K.: «Annotating emotional language data. Technical report, Verbmobil - report 236», 1999.
- FISHER, S.: «Aspectos de un itinerario. Oswald Ducrot, de la lógica a la enunciación, por el estructuralismo». En: E. NARVAJA DE ARNOUX y M^a M. GARCÍA (Eds.), *Homenaje a Oswald Ducrot*, pp. 135–145. EUDEBA, Buenos Aires, 2004.
- FÉLIX-BRASDEFER, J. C.: «Pragmatic and textual functions of *o sea*: Evidence from MexicanSpanish». En: T FACE y C. KLEE (Eds.), *Selected Proceedings of the 8th Hispanic Linguistics Symposium*, pp. 191–203. Cascadilla Proceedings Project (Available on line at: <http://www.lingref.com/cpp/hls/8/index.html>), Somerville, 2006.
- FUENTES, C.: *Introducción a la Pragmática Lingüística*. Kronos, Sevilla, 1997a.
- : «La Pragmática hoy en España: su importancia en el análisis lingüístico». En: C. FUENTES (Ed.), *Introducción Teórica a la Pragmática Lingüística*, pp. 9–26. Kronos, Sevilla, 1997b.
- : *La Sintaxis de los Relacionantes Supraoracionales*. Arco/Libros, Madrid, 1998.
- FUENTES, C. y ALCAIDE, E.R.: *Mecanismos lingüísticos de la persuasión*. Arco/Libros, Madrid, 2002.

- FULLER, J. M.: «Discourse marker use across speech contexts: a comparison of native and non-native speaker performance». *Multilingua*, 2003, **22**, pp. 185–208.
- GADNER, H.: *La nueva ciencia de la mente. Historia de la revolución cognitiva*. Paidós Transiciones, Barcelona, 1985.
- GALLARDO, B.: «En torno a la preferencia como concepto del análisis conversacional». En: E. RODRÍGUEZ y A. LÓPEZ (Eds.), *Homenaje a Enrique García*, pp. 341–354. Universidad de Valencia, Valencia, 1991.
- : «La enseñanza del componente pragmático». *Español Actual*, 1997, **67**, pp. 85–94.
- GARCÍA, J.: *Desarrollo y conocimiento*. Siglo XXI de España editores, Madrid, 1991.
- GÜELL, I.: *El cerebro al descubierto. De la emoción a la palabra*. Kairós, Barcelona, 2006.
- GENEREUX, M. y SANTINI, M.: «Exploring the use of linguistic features in sentiment analysis». *Corpus Linguistics*, 27-30 Jul, Birmingham, UK, 2007.
- GILLE, J.: *Pautas argumentativas en el diálogo espontáneo. Un estudio de conversaciones interculturales*. Tesis doctoral, Universidad de Estocolmo, 2001.
- GIVON, T.: «Toward a neuro-cognitive interpretation of context». *Pragmatics and Cognition*, 2001, **9**(2), pp. 175–201.
- GÜLICH, E. y KOTSCHI, T.: «Les marqueurs de la réformulation paraphrastique». *Cahiers de Linguistique Française*, 1983, **5**, pp. 305–35.
- GOFFMAN, E.: *La presentación de la persona en la vida cotidiana*. Amorrortu-Munguía, Madrid, 1987^a edición, 1959.
- GONZÁLEZ, E.: «Desde mi punto de vista / en mi opinión: dos signos complejos que marcan el enunciado en el que inciden como una opinión personal del locutor». *Actas del XXII Congreso Internacional de la Asociación de Jóvenes Lingüistas*. Sevilla, 2007.
- GONZÁLEZ, N.: «El material didáctico en la enseñanza del español como lengua extranjera». En: M. VILLAYANDRE (Ed.), *Actas del V Congreso de Lingüística General : León 5-8 de marzo*, volumen 2, pp. 1399–1414, 2004.
- GONZÁLEZ-LEDESMA, A.: «Pragmertext, Annotating the Spanish C-ORAL-ROM Corpus with Pragmatic Knowledge». *Proceedings of 4th Corpus Linguistics Conference, University of Birmingham, 27-30 July, 2007a*.

- GONZÁLEZ-LEDESMA, A.: «Reformulación y atenuación: los operadores de modalización *o sea* y *bueno* en el corpus oral del español C-ORAL-ROM». En: C. NICOLÁS (Ed.), *Ricerche sul Corpus del parlato romanzo C-ORAL-ROM. Studi linguistici e applicazioni didattiche per l'insegnamento di L2*, Firenze University Press, Florencia, 2007b.
- GONZÁLEZ-LEDESMA, A. y GARROTE, M.: «Los marcadores discursivos en CHIEDE, un corpus de habla infantil espontánea». *Interlingüística*, 2007, **18**.
- GONZÁLEZ-LEDESMA, A. y MORENO, A.: «Pragmatic analysis of man-machine interactions in a spontaneous speech corpus». En: *Corpus Linguistics, Birmingham*, , 2005.
- GONZÁLEZ-LEDESMA, A. y SAMY, D.: «Marcadores discursivos en árabe y en español: un estudio computacional basado en corpus paralelos con anotación pragmática». *Actas del VIII Congreso de Lingüística General: El valor de la diversidad [meta]lingüística, Madrid.*, 2008.
- GRICE, H. P.: «Lógica y conversación». En: L. M. VILLANUEVA (Ed.), *La búsqueda del significado*, pp. 511–530. Tecnos, Madrid, 1975.
- GRICE, H.P.: «Las intenciones y el significado del hablante». En: L. M. Valdes (Ed.), *La búsqueda del significado. Lecturas de filosofía del lenguaje*, pp. 511–530. Tecnos, Madrid, 1991.
- HALLIDAY, M. A. K.: *Computational and Quantitative Studies*. volumen 6. Continuum International Publishing Group, Londres, 2004.
- HAVERKATE, H.: *La cortesía verbal: estudio pragmlingüístico*. Gredos, Madrid, 1994.
- : «Estrategias de cortesía: análisis intercultural». En: M. A. CELIS y J. R. HEREDIA (Eds.), *Lengua y cultura en la enseñanza del español a extranjeros : actas del VII Congreso de ASELE*, pp. 45–56, 1998.
- HUTCHINSON, B.: *Automatic classification of discourse markers on the basis of their co-occurrences*. Tesis doctoral, University of Edinburgh, 2005.
- IGLESIAS, S.: «Los estudios de cortesía en el mundo hispánico. Estado de la cuestión.» *Oralia*, 2001, **4**, pp. 245–298.
- ISRAEL, D. A.: *El canon emocional en el sistema literario argentino*. Tesis doctoral, Universidad Autónoma de Madrid, Madrid, 2006.

- JAKOBSON, R.: «Shifters, verbal categories y the russian verb». En: L. WAUGH y M. MONVILLE-BURSTON (Eds.), *On Language: Roman Jakobson*, pp. 386–392. Harvard University Russian Language, Cambridge, 1990 (1957).
- JAYEZ, J. y ROSSARI, C.: «Discourse relations vs. discourse marker relations». *ACL'98 Workshop on Discourse Relations and Discourse Markers*, 1998, pp. 72–78.
- JOHNSON-LAIRD, P. N.: *El ordenador y la mente. Introducción a la ciencia cognitiva*. Paidós, Barcelona, 1993.
- JURAFSKY, D.: «Pragmatics and Computational Linguistics». En: L. R. HORN y G. WARD (Eds.), *Handbook of Pragmatics*, Blackwell, Oxford, 1995.
- KARTTUNEN, L.: «Presuposición y contexto lingüístico». En: M. T. JULIO y R. MUÑOZ 1998 (Eds.), *Textos clásicos de pragmática*, pp. 175–192. Arco/Libros, Madrid, 1977.
- KIRK, J.: «Integrating Prosody, Pragmatics and Syntax in a Corpus-based Linguistic Description of Irish Standard English». *Proyecto de Investigación de la Universidad de Queen's. Belfast. 2003-2005*, 2005.
- KOSTOULAS, T. ; MPORAS I., T.; GANCHEV y FAKOTAKIS, N.: «A Real-World Emotional Speech Corpus for Modern Greek». En: European Language Resources Association (ELRA) (Ed.), *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, 2008.
- LAHUERTA, A. C. y PELAYO, M. F.: «Usos marginales de los marcadores del discurso y su efecto en la comprensión lectora del español como lengua extranjera». *Ibérica*, 2003, 5, pp. 49–68.
- LAKOFF, G. y JOHNSON, M.: *Metáforas de la vida cotidiana*. Cátedra, Madrid, 1980.
- LAKOFF, R.: «La lógica de la cortesía o acuérdate de dar las gracias». En: M. T. & MUÑOZ R. JULIO (Ed.), *Textos clásicos de pragmática*, pp. 259–278. Arco Libros, Madrid, 1998.
- LAMIROY, B.: «Pragmatic connectives and L2 acquisition: the case of French and Dutch». *Journal of Pragmatics*, 1994, 4:2, pp. 183–201.

- LASKOWSKI, K. y BURGER, S.: «Development of an Annotation Scheme for Emotionally Relevant Behavior in Multiparty Meeting Speech». *The 9th European Conference on Speech Communication and Technology (INTERSPEECH2005), Lisbon, Portugal, 04-08 September, 2005*.
- : «Annotation and Analysis of Emotionally Relevant Behavior in the ISL Meeting Corpus». En: *Proceedings of the 5th ELRA International Conference on Language Resources and Evaluation, LREC, Genoa, Italy, 26 Mai, , 2006*.
- LEGRENZI, P.: 1998). *Cómo funciona la mente*. Psicología Alianza Editorial, Madrid, 2000, (1ª Ed. 1998).
- LEÓN, J.A. (Ed.): *Conocimiento y discurso. Claves para inferir y comprender*. Pirámide, Madrid, 2003.
- LEVINSON, S.: *Pragmática*. Teide, Barcelona, 1989.
- LÓPEZ, H.: *Metodología de la investigación lingüística*. Editorial Colegio España, Salamanca, 1994.
- LUQUE, J. D. y MANJÓN, F.J.: «Claves culturales del diseño de las lenguas: fundamentos de tipología fraseológica». *Estudios de Lingüística Española*, 2002, **16**.
- MAFFIA, D.: «Conocimiento y emoción». *Arbor: Ciencia, pensamiento y cultura. Ejemplar dedicado a: Ciencia, tecnología y valores desde una perspectiva de género*, 2005, **716**, pp. 515–521.
- MARCHANTE, M^a. P.: «El tratamiento de los marcadores contraargumentativos en los manuales de E/LE». *Biblioteca virtual Redele de la Consejería de Educación en Australia y Nueva Zelanda*, 2005.
<http://www.mepsyd.es/redele/biblioteca2005/fiape/marchante.pdf>
- MARCONDE, D.: «Desfazendo mitos sobre a pragmática». *Alceu*, 2000, **1(1)**, pp. 38–46.
- MARCU, D.: «The rhetorical parsing of natural language texts». *Proceedings of the 35th annual meeting on Association for Computational Linguistics, Madrid, Spain, 1997*, pp. 96–103.
- : «A Surface-Based Approach To Identifying Discourse Markers And Elementary Textual Units In Unrestricted Texts». *Workshop On Discourse Relations And Discourse Markers*, 1998.

- MARINA, J. A.: *Teoría de la Inteligencia Creadora*. Anagrama, Barcelona, 1993.
- MARÍN, F. M.: «La comunicación mediante la computadora y la enseñanza de segundas lenguas». *II Congreso Internacional de la ACH, Asociación Coreana de Hispanistas, 28-29 de junio, Alcalá de Henares (Madrid), 2002*.
- MARTIN, J. C.; CARIDAKIS, G.; DEVILLER, L.; KARPOUZIS, K. y ABRILIAN, S.: «Manual Annotation and Image Processing of Multimodal Emotional Behaviours in TV Interviews». En: I. MAGLOGIANNIS; K. KARPOUZIS y M. BRAMER (Eds.), *Artificial Intelligence Applications and Innovations*, volumen 1127-1132, pp. 369–377. Springer, Boston, 2006.
- MARTÍN, M. y MONTOLÍO, E. (Eds.): *Los marcadores del discurso. Teoría y análisis*. Arco/Libros, Madrid, 1998.
- MARTÍN, M^a. A. y PORTOLÉS, J.: «Los marcadores del discurso». En: V. DEMONTE y I. BOSQUE (Eds.), *Gramática Descriptiva*, Espasa-Calpe, Madrid, 1999.
- MATTE, F.: *Gramática comunicativa del español: de la idea a la lengua*. volumen 2. Difusión, Madrid, 1992.
- MISHKOFF, H. C.: *A fondo: Inteligencia Artificial*. Anaya, Madrid, 1988.
- MONDADA, L.: «Por una Lingüística Interaccional». *Discurso y Sociedad*, 2001, **3**, pp. 61–89.
- MONTOLÍO, E.: «Si me lo permiten...: Gramática y pragmática: sobre algunas estructuras condicionales regulativas en español». *Diálogos hispánicos de Amsterdam (Ejemplar dedicado a: Aproximaciones pragmalingüísticas al español)*, 1993, **12**, pp. 119–148.
- : «La teoría de la relevancia y el estudio de los marcadores del discurso». En: *Los marcadores del discurso. Teoría y Análisis*, pp. 93–121. Arco/Libros, Madrid, 1998.
- : «La interacción entre gramática y pragmática en el aula de ELE». En: *XIII Encuentro Práctico de Profesores de ELE organizado por International House Barcelona y Difusión, Centro de Investigación y Publicaciones de Idiomas*, , 2004.
- MONTOLÍO, E. y UNAMUNO, V.: «El marcador del discurso *a ver* (catalan *a veure*) en la interacción profesor-alumno». En: J. BUSTOS (Ed.), *Lengua, discurso, texto: I Simposio internacional de análisis del discurso*, pp. 603–620. Arco/Libros, Madrid, 2000.

- MORA, F.: *El reloj de la sabiduría. Tiempos y espacios en el cerebro humano*. Biología. Alianza Editorial, Madrid, 2001.
- MORENO, A.: *La Lingüística Computacional*. Síntesis, Madrid, 1998.
- : «La evolución de los corpus de habla espontánea: la experiencia del LLI-UAM». *Actas de las Segundas Jornadas de Tecnologías del Habla, Granada. Diciembre, 2002a*.
- : «Los corpus orales del LLI-UAM: primera generación y segunda generación». *La Musa Digital (Revista Electrónica)*, 2002b.
- MURILLO, S.: «Teaching discourse markers: *On the other hand, on the contrary* and their Spanish equivalents». En: I. VÁZQUEZ y I. GUILLÉN (Eds.), *Los distintos dominios de la Lingüística Aplicada desde la perspectiva de la Pragmática*. Textos de Filología 6, pp. 313–318. Anubar, Zaragoza, 1998.
- MURTAGH, T.: *El cosmos invisible. Luz fantástica. Enciclopedia de la Astronomía y el espacio*. Discovery Chanel, 2004.
- NAVARRO, B.; CIVIT, M.; MARTÍN, M^a A.; MARCOS, R. y FERNÁNDEZ, B.: «Syntactic, semantic and pragmatic annotation in Cast3LB». En: *Corpus Linguistics 2003 Workshop on Shallow Procesing of Large Corpora. UCREL Technical Report, Lancaster (UK), marzo, , 2003*.
- NEBRIJA, EQUIPO PEDAGÓGICO: *Historia de la metodología de las lenguas extranjeras*. Fundación Antonio de Nebrija, Madrid, 1997.
- : *Principios metodológicos de los enfoques comunicativos*. Fundación Antonio de Nebrija, Madrid, 1998.
- NIETO, J. M.: *Introducción al análisis del discurso hablado*. Universidad de Granada, Granada, 1995.
- NISSAN, E.: «Emotion, culture, communication». *Pragmatics and cognition*, 1997, **5(2)**, p. 355 369.
- NURMI, A.: «Employing and elaborating annotation for the study of modality». *Studies in Variation, Contacts and Change in English 1: Annotating Variation and Change*, 2007.
<http://www.helsinki.fi/varieng/journal/volumes/01/nurmi/>
- OCHS, E.: «La universalidad de los postulados conversacionales». En: M. T. JULIO y R. MUÑOZ 1998 (Eds.), *Textos clásicos de pragmática*, Arco/Libros, Madrid, 1977.

- O'DONELL, M.: «RST ANNOTATION TOOL». <http://www.isi.edu/licensed-sw/RSTTool/index.html>, 1997point
- ONG, W.: *Oralidad y escritura: tecnologías de la palabra*. Fondo de Cultura Económica de México, Buenos Aires, 1993.
- PAGGIO, P.: «Annotating Information Structure in a Corpus of Spoken Danish». En: N. CALZOLARI; K. CHOUKRI; B. MAEGAARD; J. MARIANI; J. ODJIK; S. PIPERIDIS y D. TAPIAS (Eds.), *Proceedings from the 5th International Conference on Language Resources and Evaluation, Genova 24-24, May, , 2006*.
- PENADÉS, I.: «Aproximación pragmática a las unidades fraseológicas». *Homenaje al profesor A. Roldán Pérez*, 1997, **1**, pp. 411–426.
- PINKER, S.: *¿Cómo funciona la mente?* Destino, Barcelona, 1997.
- POESIO, M. y ARSTEIN, R.: «Anaphoric Annotation in the ARRAU Corpus». En: *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08). European Language Resources Association (ELRA)*, Marrakech, Marruecos, 2008.
- POPESCU-BELIS, A. y ZUFFEREY, S.: «Automatic Identification of Discourse Markers in Multiparty Dialogues». *Working Paper 65*, 2006.
- PORTOLÉS, J.: «Inferencias axiológicas convencionales en español». *Lenguajes naturales y lenguajes formales*, 1992, **VIII**, pp. 515–522.
- : «La distinción entre los conectores y otros marcadores del discurso en español». *Verba*, 1993, **20**, pp. 141–170.
- : «Pertinencia y pragmática». *Revista de Occidente*, 1994, **154**, pp. 55–66.
- : «La teoría de la argumentación en la lengua y los marcadores del discurso». En: M. MARTÍN y E. DURÁN (Eds.), *Los marcadores del discurso. Teoría y análisis*, pp. 71–93. Arco/Libros, Madrid, 1998a.
- : «La teoría de la argumentación en la lengua y los marcadores del discurso». En: M. MARTÍN y E. DURÁN (Eds.), *Los marcadores del discurso. Teoría y análisis*, pp. 71–93. Arco/Libros, Madrid, 1998b.
- : *Marcadores del discurso*. Ariel, Barcelona, 1998c.
- : *Pragmática para hispanistas*. Síntesis, Madrid, 2004.

- : «La escritura y los marcadores del discurso». En: J. A. MOYA (Ed.), *Pragmática y enseñanza de la lengua española. Actas de las X Jornadas sobre la enseñanza de la lengua española*, pp. 37–53. Universidad de Granada, Granada, 2005.
- PRADA, J. J. y MONCECCHI, G.: «Reconocimiento eficiente de marcadores del discurso en español». *VIII Simposio Internacional de Comunicación Social, Santiago de Cuba, Cuba, January*, 2003.
- PUNSET, E.: *El alma está en el cerebro. Radiografía de la máquina de pensar*. Aguilar, Madrid, 2006.
- RATEY, J.J.: *El cerebro: manual de instrucciones*. Mondadori, Barcelona, 2002.
- RÁBADE, S.: *Teoría del conocimiento*. Akal, Madrid, 1995.
- REAL ACADEMIA ESPAÑOLA, X: *Diccionario de la lengua española*. Espasa Calpe, Madrid, 21ª edición, 2000.
- REYES, G.: *La pragmática lingüística. El estudio del uso del lenguaje*. Montesinos, Barcelona, 1990.
- RODRÍGUEZ, L.J.: *Estudio y modelización acústica del habla espontánea en diálogos hombre-máquina y entre personas*. Tesis doctoral, Universidad del País Vasco, 2004.
- RODRÍGUEZ, L.J.; TORRES, I. y VARONA, A.: «Annotation and analysis of disfluencies in a spontaneous speech corpus in Spanish». *Disfluency in Spontaneous Speech (DiSS'01), ISCA Tutorial and Research Workshop (ITRW), Edinburgh, Scotland, UK, August 29-31*, 2001.
- RODRÍGUEZ MARTÍN, J. R.: «El uso de Internet en el aula de ELE». *redELE*, 2004, **2**.
<http://www.sgci.mec.es/redele/revista2/rodriguez.shtml>
- RODRIGNO, L.; GARCÍA, A. y MARTÍNEZ, P.: «Planteamiento semántico y pragmático para gestión de diálogos en asistentes virtuales». *Procesamiento del lenguaje natural*, 2002, **28**, pp. 81–88.
- SAMY, D. y GONZÁLEZ-LEDESMA, A.: «Pragmatic Annotation of Discourse Markers in a Multilingual Parallel Corpus (Arabic- Spanish-English)». *Proceedings of LREC 2008, Marrakech, 28-30 de mayo*, 2008.
- SAYGIN, A. P. y CICEKLI, I.: «Pragmatics in human-computer conversations». *Journal of Pragmatics*, 2002, **34**, pp. 227–258.

- SCHUMANN, J. H.: «A neurobiological basis for decision making in language pragmatics». *Pragmatics & Cognition*, 1999, **7(2)**, pp. 283–311.
- SERRANO, S.: *La Lingüística: su historia y su desarrollo*. Biblioteca de divulgación temática, Barcelona, 1983.
- SIMONS, G. L.: *Introducción a la inteligencia artificial*. Díaz de Santos, Madrid, 1988.
- SÁNCHEZ, F.: *Corpus de Referencia del Español Actual CREA y corpus Diacrónico del Español (CORDE) . Corpus y paradigma verbal*. Universidad de Salamanca, Salamanca, 1998.
- SORIA, C.; CATTONI, R. y DANIELI, M.: «ADAM: An architecture for xml-based dialogue annotation on multiple levels». En: L. DYBKJAER; K. HASIDA y D. TRAUM (Eds.), *Proceedings of the First SIGDial Workshop on Discourse and Dialogue*, pp. 9–18. Association for Computational Linguistics, Hong Kong, 2000.
- SPERBER, D. y WILSON, D.: *La Relevancia: comunicación y procesos cognitivos*. Visor, Madrid, 1994.
- STEDE, M.; WANNER, L. y HOVY, E.: «Discourse Relations and Discourse Markers». *Proceedings of the Workshop Sponsored by The Association for Computational Linguistics COLING-ACL '98*, 1998, pp. 36–42.
- STOYANOV, V. y CARDIE, C.: «Annotating Topics of Opinions». En: *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA), Marrakech, Marruecos, 2008.
- TABOADA, M.: «Discourse Markers as signal (or not) of rethorical relations». *Journal of Pragmatics*, 2006, **38**, pp. 567–592.
- TAULÉ, M.; MARTÍ, M^a A. y RECASENS, M.: «AnCora: Multilevel Annotated Corpora for Catalan and Spanish». En: European Language Resources Association (ELRA) (Ed.), *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Marruecos, 2008.
- TESTON, S. y VERÓNIS, J.: «Recherche de critères formels pour l'identification automatique des particules discursives». *Journée d'étude de l'ATALA Modéliser et décrire l'organisation discursive à l'heure du document numérique*. La Rochelle, <http://www.up.univ-mrs/pdf/2004-LaRochelle-Part-Disc.pdf>, 2004.

- TEUFEL, S y MOENS, M. F.: «Discourse-level argumentation in scientific articles: human and automatic annotation». En: M. WALKER (Ed.), *Towards Standards and Tools for Discourse Tagging: Proceedings of the Workshop*, pp. 84–93. Association for Computational Linguistics, Somerset, New Jersey, 1999.
- TORDESILLAS, M. y NEGRONI, M.: *La enunciación en la lengua. De la deixis a la polifonía*. Gredos, Madrid, 2001.
- TORRES, M^a A.: *La interjección*. Universidad de Cádiz, Cádiz, 2000.
- T'SOU, B.K; LAI, B.Y. T; CHAN, S.W.K; GAO, W. y ZHAN, X.: «Enhancement of a Chinese Discourse Marker Tagger with C4.5». <http://citeseer.ist.psu.edu/572778.html>, 2000.
- TUSÓN, A.: *Análisis conversacional*. Ariel Practicum, Barcelona, 1997.
- VIAGARA, A. M.: «La función fática del lenguaje con especial atención a la lengua hablada». *Actas del congreso de la sociedad española de lingüística*, 1990, **2**, pp. 1088–1097.
- WEISSER, M.: «Tagging Dialogues in SPAACy». *Traitement Automatique des Langues: Le traitement automatique des corpus oraux*, 2004, **45(2)**, pp. 131–157.
- WICHMANN, A. y BLAKEMORE, D.: «The prosody-pragmatics interface». *Journal of Pragmatics*, 2006, **38**, pp. 1537–1541.
- WIEBE, J.; WILSON, T.; BRUCE, R.; BELL, M. y MARTIN, M.: «Learning Subjective Language». *Computational Linguistics*, 2004, **30(3)**, pp. 277–308.
- WIKINS, D.P.: «Interjection as deictics». *Journal of Pragmatics*, 1992, **18**, pp. 119–158.
- WILSON, T.: «Annotating Subjective Content in Meetings». En: *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08). European Language Resources Association (ELRA)*, Marrakech, Marruecos, 2008.
- WOLF, M.: *Sociologías de la vida cotidiana*. Cátedra, Madrid, 1979.
- YOUNG, J.Z.: *Antropología Física: Introducción al estudio del hombre*. Vicens Vives, Barcelona, 1976.
- ZAVADIVKER, M. N.: «La metáfora como recurso epistémico». *A Parte Rei, Revista de Filosofía*, 2005, **40**, pp. http://serbal.pntic_mec.es/cmuno11/index.html.

ZUFFEREY, S. y POPESCU-BELIS, A.: «Towards Automatic Identification of Discourse Markers in Dialogs: The Case of Like». *SIGDial '04*, 2004, pp. 63–71.